

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2002, 07:34:31 : Search time 26 Seconds
(without alignments)
194.620 Million cell updates/sec

Title: US-09-733-685-2

Perfect score: 623

Sequence: 1 MNLSLKEERVEDNCKSG.....QGLGLDNCKPEPDSVLSL 122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result, being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80.5	12.9	968	PKD2_HUMAN	Q13563 homo sapien
2	77	12.4	966	PKD2_MOUSE	Q35245 mus musculus
3	76	12.2	1007	Y74_CHLMO	Q9P166 chlamydia m
4	72.5	11.6	448	YNQ6_YEAST	P53890 saccharomyc
5	72.5	11.6	493	GALT_STRSL	Q8V592 streptococc
6	72	11.6	444	CHLB_CHLPT	P37824 chlamydomon
7	71.5	11.5	543	THSB_TREAC	P48425 thermoplasm
8	71	11.4	217	GRE1_HUMAN	Q9HAV7 homo sapien
9	71	11.4	879	RA50_SULIC	Q96VR5 scilfolobus
10	70.5	11.3	301	TRT2_CHICK	P02642 gallus gall
11	70.5	11.3	775	RIR1_VZVD	P09248 varicella-z
12	70	11.2	320	IF34_HUMAN	Q75821 homo sapien
13	69	11.1	149	SP17_MOUSE	Q62252 mus musculus
14	69	11.1	843	CYP1_BRUMA	Q27450 bugia mala
15	69	11.1	2319	AKA6_HUMAN	Q13023 homo sapien
16	68.5	11.0	245	TRJA_THEMA	Q9X1R0 thermotoga
17	68.5	11.0	569	Y05_DROME	Q9X719 drosophila
18	68.5	11.0	837	RA53_METH	Q26640 methanobact
19	68.5	11.0	1167	CAGA_HELPJ	Q9Z1T1 helicobacte
20	68	10.9	320	IF34_MOUSE	Q921D1 mus musculus
21	68	10.9	499	UDPG_YEAST	P32861 saccharomyc
22	68	10.9	723	MY5B_MOUSE	P21271 mus musculus
23	68	10.9	1539	Y373_HUMAN	Q15078 homo sapien
24	67.5	10.8	105	BL24_THEMA	P38513 thermotoga
25	67.5	10.8	240	SNF_YEAST	P39929 saccharomyc
26	67.5	10.8	513	CATB_PSEAE	Q59635 pseudomonas
27	67.5	10.8	864	RA5C_SULSO	Q97WH0 sulfolobus
28	67	10.8	153	SCDC_DJBHA	Q42724 debaryomyce
29	67	10.8	562	CHS5_CANAL	Q74161 candida alb
30	67	10.8	596	Y41J_RHSN	P55493 rhizobium s
31	66.5	10.7	542	TULL_HUMAN	Q00294 homo sapien
32	66.5	10.7	569	YX1D_BACSU	P42295 bacillus su
33	66.5	10.7	668	YM18_YEAST	Q04511 saccharomyc

RESULT 1
PKD2_HUMAN
ID PKD2_HUMAN STANDARD: PRT: 968 AA.
AC Q13563: 060441: Q15764:
ST 15-JUL-1999 (Rel. 38, Created)
ST 15-JUL-1999 (Rel. 38, Last sequence update)
ST 16-OCT-2001 (Rel. 40, Last annotation update)
DE Polycystin 2 (Autosomal dominant polycystic kidney disease type II protein) (Polycystin) (R48321).
GN PKD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RF SEQUENCE FROM N.A.
RA Mochizuki T., Wu G., Hayashi T., Xenophontos S.L., Veldhuisen B., Saris J.J., Reynolds D.M., Cai Y., Gabow P.A., Pierides A., Kimberling W.J., Breuning M.H., Deltas C.C., Peters D.J.M., Somlo S.; *PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein.*
RT Science 272:1339-1342(1996).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-97432829; PubMed-9286709;
RA Hayashi T., Mochizuki T., Reynolds D.M., Wu G., Cai Y., Somlo S.; *Characterization of the exon structure of the polycystic kidney disease 2 gene (PKD2).*
RL Genomics 44:131-136(1997).
RN [3]
RP SEQUENCE OF 361-968 FROM N.A.
RC TISSUE=Breast;
RA Schneider M.C., Rodriguez A., Nomura H., Zhou J., Morton C.C., Reeders S.T., Wierowicz S.; *A gene similar to PKD1 maps to chromosome 4q22: a candidate gene for PKD2.*
RL Genomics 38:1-4(1996).
RN [4]
RP VARIANT ADPKD GLY 414.
RX MEDLINE-97465600; PubMed-9326320;
RA Veldhuisen B., Saris J.J., de Haij S., Hayashi T., Reynolds D.M., Mochizuki T., Elies R., Fossdal R., Bogdanova N., van Dijk M.A., Coto E., Ravine D., Noerby S., Verellen-Dumoulin C., Breuning M.H., Somlo S., Peters D.J.M.; *A spectrum of mutations in the second gene for autosomal dominant polycystic kidney disease (PKD2).*
RL Am. J. Hum. Genet. 62:547-555(1997).
RN [5]
RP VARIANT ADPKD PRO-356 AND VARIANT PRO-28.
RX MEDLINE-99340490; PubMed-10411676;
RA Tofia R., Viribay M., Telleria D., Badenas C., Watson M., Harris P.C., Darnell A., San Millan J.L.; *Seven novel mutations of the PKD2 gene in families with autosomal dominant polycystic kidney disease.*

ALIGNMENTS

34	66.5	10.7	681	1	MOGL_CAEEL	044411 caenorhabdi
35	66.5	10.7	866	1	MYSP_SCHJA	Q05670 schistosoma
36	65.5	10.5	399	1	HBAL_SCHPO	Q05146 schistosoma
37	65.5	10.5	405	1	PI52_MOUSE	P01072 mus musculus
38	65	10.4	237	1	COAT_TOBSV	P03598 tobacco str
39	65	10.4	527	1	TCPB_YEAST	P39076 saccharomyc
40	65	10.4	1040	1	EG15_CAEEL	Q10656 caenorhabdi
41	64.5	10.4	346	1	YE77_METJA	Q58872 methanococc
42	64.5	10.4	381	1	E28B_YEAST	P32502 saccharomyc
43	64.5	10.4	398	1	THIL_YEAST	P41338 saccharomyc
44	64.5	10.4	406	1	PI52_HUMAN	P48426 homo sapien
45	64.5	10.4	466	1	SYC_RHINE	Q92R20 rhizobium s

Kidney Int. 56:28-33(1999).

[6]

CC -1- FUNCTION: PKD1 AND PKD2 MAY FUNCTION THROUGH A COMMON SIGNALING PATHWAY THAT IS NECESSARY FOR NORMAL TUBULOGENESIS.

CC -1- SUBUNIT: INTERACTS WITH PKD1. PKD1 REQUIRES THE PRESENCE OF PKD2 FOR STABLE EXPRESSION.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).

CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN OVARY, FETAL AND ADULT KIDNEY, TESTIS, AND SMALL INTESTINE. NOT DETECTED IN PERIPHERAL LEUKOCYTES.

CC -1- DOMAIN: THE C-TERMINAL IS IMPLICATED IN THE INTERACTION WITH PKD1.

CC -1- DISEASE: DEFECTS IN PKD2 ARE THE CAUSE OF AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE TYPE II (ADPKD TYPE II), WHICH REPRESENTS APPROXIMATELY 15% OF CASES OF AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE (ADPKD); A COMMON AUTOSOMAL DOMINANT GENETIC DISEASE AFFECTING ABOUT 1 OUT 1000 INDIVIDUALS. IT IS CHARACTERIZED BY PROGRESSIVE FORMATION AND ENLARGEMENT OF CYSTS IN BOTH KIDNEYS, TYPICALLY LEADING TO END-STAGE RENAL DISEASE IN ADULT LIFE. CYSTS ALSO OCCURS IN THE LIVER AND OTHER ORGANS. ALL MUTATIONS, SCATTERED BETWEEN EXONS 1 AND 11, RESULT IN A TRUNCATED PKD2 THAT LACKS BOTH THE CALCIUM-BINDING EF-HAND DOMAIN AND THE TWO CYTOPLASMIC DOMAINS REQUIRED FOR THE INTERACTION OF PKD2 WITH PKD1 AND WITH ITSELF. PKD2 IS CLINICALLY Milder THAN PKD1, BUT IT HAS A DELETERIOUS IMPACT ON OVERALL LIFE EXPECTANCY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

DR EMBL; U50928; AAC50520.1; .

DR EMBL; AF004873; AAC16004.1; JOINED.

DR EMBL; AF004859; AAC16004.1; JOINED.

DR EMBL; AF004860; AAC16004.1; JOINED.

DR EMBL; AF004861; AAC16004.1; JOINED.

DR EMBL; AF004862; AAC16004.1; JOINED.

DR EMBL; AF004863; AAC16004.1; JOINED.

DR EMBL; AF004864; AAC16004.1; JOINED.

DR EMBL; AF004865; AAC16004.1; JOINED.

DR EMBL; AF004866; AAC16004.1; JOINED.

DR EMBL; AF004867; AAC16004.1; JOINED.

DR EMBL; AF004868; AAC16004.1; JOINED.

DR EMBL; AF004869; AAC16004.1; JOINED.

DR EMBL; AF004870; AAC16004.1; JOINED.

DR EMBL; AF004871; AAC16004.1; JOINED.

DR EMBL; AF004872; AAC16004.1; JOINED.

DR EMBL; U56813; AAC50333.1; .

DR Genbank; HGNC:9009; PKD2.

DR MIM; 173910; .

DR InterPro; IPR001682; Ca/Na_pore.

DR InterPro; IPR002111; Cal_channel_TripL.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000636; M-channel_nig.

DR InterPro; IPR003915; PKD_2.

DR Pfam; PF00520; Ion_trans_1.

DR PRINTS; PR01433; POLYCYSTIN2.

DR PROSITE; PS00018; EF_HAND; FALSE_NEG.

KW Transmembrane; Glycoprotein; Coiled coil; Calcium-binding; Disease mutation; Polymorphism.

FT DOMAIN 1 223 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 224 244 POTENTIAL.

FT DOMAIN 245 468 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 469 489 POTENTIAL.

FT

110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
MEDLINE=98384541; PubMed=9716561;
Pennekamp P., Boydanova N., Wilda M., Markoff A., Hameister H.,
Forst J., Dworniczak B.,
"Characterization of the murine polycystic kidney disease (Pkd2)
gene".
Mamm. Genome 9:749-752(1998).
-!- FUNCTION: PKD1 AND PKD2 MAY FUNCTION THROUGH A COMMON SIGNALING
PATHWAY THAT IS NECESSARY FOR NORMAL TUBULOGENESIS (BY
SIMILARITY).
-!- SUBUNIT: INTERACTS WITH PKD1. PKD1 REQUIRES THE PRESENCE OF PKD2
FOR STABLE EXPRESSION (BY SIMILARITY).
-!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
-!- TISSUE SPECIFICITY: EXPRESSED IN MESENCHYMALLY DERIVED STRUCTURES
IN THE DEVELOPING EMBRYO AT DAY 12.5. IN ADULT, MOSTLY EXPRESSED
IN KIDNEY.
-!- DOMAIN: THE C-TERMINAL IS IMPLICATED IN THE INTERACTION WITH PKD1
(BY SIMILARITY).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: AF014010; AAC53388.1; -
EMBL: Y13278; CAA74521.1; -
EMBL: Y14105; CAA74551.1; -
EMBL: Y14106; CAA74551.1; JOINED.
EMBL: Y14107; CAA74551.1; JOINED.
EMBL: Y14108; CAA74551.1; JOINED.
EMBL: Y14109; CAA74551.1; JOINED.
EMBL: Y14110; CAA74551.1; JOINED.
EMBL: Y14111; CAA74551.1; JOINED.
EMBL: Y14112; CAA74551.1; JOINED.
EMBL: Y14113; CAA74551.1; JOINED.
EMBL: Y14114; CAA74551.1; JOINED.
EMBL: Y14115; CAA74551.1; JOINED.
EMBL: Y14116; CAA74551.1; JOINED.
EMBL: Y14117; CAA74551.1; JOINED.
EMBL: Y14118; CAA74551.1; JOINED.
EMBL: Y14119; CAA74551.1; JOINED.
EMBL: Y14120; CAA74552.1; -
HSP: P02632; ICBI.
MGD: MGI:1059818; Pkd2.
InterPro: IPR001682; Ca/Na_poro.
InterPro: IPR002111; Cat_channel_TrpL.
InterPro: IPR002048; EF-hand.
InterPro: IPR000636; M+channel_nig.
InterPro: IPR003915; PKD_2.
Pfam: PF00520; ion_trans...
PRINTS: PR01433; polycystin2.
PROSITE: PS00018; EF_HAND; FALSE_NEG.
Transmembrane: Glycoprotein; Coiled coil; Calcium-binding.
CYTOPLASMIC (POTENTIAL).
DOMAIN 1 221
TRANSMEM 222 242
DOMAIN 243 466
TRANSMEM 467 487
DOMAIN 488 503
TRANSMEM 504 524
DOMAIN 525 548
TRANSMEM 549 569
DOMAIN 570 596
TRANSMEM 597 617
DOMAIN 618 656
TRANSMEM 657 677
DOMAIN 678 966
TRANSMEM 95 99
DOMAIN 151 154
TRANSMEM 761 772
CA_BIND 772 794
DOMAIN 761 794
COILED COIL (POTENTIAL).

FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 365 365 I -> M (IN REF. 2).
FT CONFLICT 370 370 R -> K (IN REF. 2).
FT CONFLICT 560 560 A -> S (IN REF. 2).
FT CONFLICT 688 688 DL -> SV (IN REF. 2; CAA74551).
FT CONFLICT 746 746 K -> E (IN REF. 2).
FT CONFLICT 942 942 S -> P (IN REF. 2).
FT CONFLICT 957 957 S -> G (IN REF. 2).
SQ SEQUENCE 966 AA: 108995 MW: 8888889049935F CRC64;
Query Match 12.4%; Score 77; DB 1: Length 966;
Best Local Similarity 29.1%; Pred. No. 12;
Matches 30; Conservative 17; Mismatches 44; Indels 12; Gaps 3;
QY 8 FERVEEDNKGSDNRRGKPTSTVVRIVTEEEVDEFFKILRRV----HVATRIYAKVNGVA 63
DY 811 FEUDDSDGSHSRHRSISGV-----SYEEFQVLRVVRDMEHSIGSIYSKIDAVIV 86;
QY 64 EDELPRK-RRKRSNGLIKNSLDCNGVRDGEFDEINRVGLQGL 105
DY 664 KLEIMPRAKLKREVLGRLLDGVAEADARLGRDSEIHRQMERL 906

RESULT 3
Y741_CHLMU STANDARD: PRI: 1007 AA.
AC Q9PJT6;
DI 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TC0741 precursor.
GN TC0741.
OS Chlamydia muridarum.
OX Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83560;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
Linher K., Weidman J., Kouri H., Craven B., Bowman C., Dodson R.,
Swinn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
Nucleic Acids Res. 28:1397-1406(2000).
RT -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0572/CT456/TC0741
FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE002342; AAF39550.1; -
DR TIGR: TC0741; -
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 51 POTENTIAL.
FT CHAIN 52 1007 HYPOTHETICAL PROTEIN TC0741.
SQ SEQUENCE 1007 AA: 104906 MW: 842800C0871E1518 CRC64;
Query Match 12.2%; Score 76; DB 1: Length 1007;
Best Local Similarity 32.2%; Pred. No. 15;
Matches 29; Conservative 11; Mismatches 32; Indels 18; Gaps 5;

CC -!- PATHWAY: Light-independent chlorophyll biosynthesis.
CC -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
CC of three subunits: chl_a, chl_b and chl_c. Could form a
CC heterotetramer of two chl_a and two chl_b subunits (by similarity).
CC -!- SIMILARITY: BELONGS TO THE CHL_a / BCH₂ / BCHZ FAMILY.
DR InterPro: IPR000510; Oxred_nitrogsnel.
DR Pfam: PF00148; oxidored_nitro_1.
DR Oxidoreductase; Photosynthesis; Chlorophyll biosynthesis; Chloroplast.
FT NON_TER 444 444
FT SEQUENCE 444 AA: 50058 MW: 650652CAAE4857E CRC54;
Query Match 11.68; Score 72; DB 1; Length 444;
Best Local Similarity 25.08; Pred. No. 14;
Matches 32; Conservative 17; Mismatches 44; Indels 30; Gaps 5;
DY 11 VEDNGKSDGNGKSGTEVVRTVTEEEEDFEFKILRRVHVHVAITRVAKVNGVAGE--- 56
b 113 ISESNHGSGIDKATSDVILADVNNHYRNE-----LQADRTLEGIVRFYLEKEKLN 165
y 57 ---LPSKKRRKRSQN-----LGLRNSLDCNGVRGDEFEINRVGLQGLGLEN-CKPEPD 116
b 166 INTIPTK-KKPSANTIGFTILGFHMHQDCR-----ELKRL-ENNLG-EVNEIIPGG 216
DY 117 SVS 119
b 217 SVT 219
RESULT 7
HSB_THEAC
D THSB THEAC STANDARD; PRT: 543 AA.
C P48425;
T 01-FEB-1996 (Rel. 33, Created)
T 01-FEB-1996 (Rel. 33, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Thermosome beta subunit (Thermosome subunit 2) (Chaperonin beta
E subunit).
N THSB OR TA1276.
S Thermoplasma acidophilum.
S Archaea; Euryarchaeota; Thermoplasmatia; Thermoplasmatiales;
C Thermoplasmataceae; Thermoplasma.
X NCBI_TaxId=2303;
P [1]
P SEQUENCE FROM N.A., AND SEQUENCE OF 188-195.
C STRAIN-ATCC 25905;
X MEDLINE=95314774; PubMed=7794526;
X Waldmann T., Lupas A.N., Kellermann J., Peters J., Baumeister W.;
T "Primary structure of the thermosome from Thermoplasma acidophilum.";
T Biol. Chem. Hoppe-Seyler 376:119-126(1995).
P [2]
P SEQUENCE FROM N.A.
C STRAIN-DSM 1728;
X MEDLINE=20479972; PubMed=11029001;
X Ruepp A., Granel W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
X Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
T "The genome sequence of the thermoacidophilic scavenger Thermoplasma
T acidophilum.";
T Nature 407:508-513(2000).
P [3]
P SEQUENCE OF 80-113 AND 421-445.
C STRAIN-ATCC 25905;
X MEDLINE=95172073; PubMed=7867646;
X Waldmann T., Nimesgern E., Nitsch M., Peters J., Pfeifer G.,
X Mueller S., Kellermann J., Engel A., Hartl F.-O., Baumeister W.;
T "The thermosome of Thermoplasma acidophilum and its relationship to
T the eukaryotic chaperonin Tric.";
T Eur. J. Biochem. 227:848-856(1995).
P [4]
P X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).
X MEDLINE=98206474; PubMed=9546398;
A Ditzel L., Lowe J., Stock D., Stetter K.O., Huber H., Huber R.,
A Steinbacher S.;
T "Crystal structure of the thermosome, the archaeal chaperonin; and

RT HOMOLOG OF OCT.";
RL Cell 93:125-138(1998).
CC -!- FUNCTION: MOLECULAR CHAPERONE; BINDS UNFOLDED POLYPEPTIDES IN
CC V-TRO. AND HAS A WEAK ATPASE ACTIVITY.
CC -!- SUBUNIT: FORMS A HETERO-OLIGOMERIC COMPLEX OF TWO STACKED EIGHT-
CC MEMBERED RINGS.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: Z46650; CA886611.1;
DR EMBL: A2445067; CAC22400.1;
DR PDB: 1A6E; 23-MAR-99.
DR PUB: 1A6E; 23-MAR-99.
DR InterPro: IPR002194; Chaperonin_TCP-1.
DR InterPro: IPR002423; Cpn60_TCP-1.
DR Pfam: PF00118; cpn60_TCP1_1.
DR PROSITE: PS00750; TCP1_1.
DR PROSITE: PS00751; TCP1_2.
DR PROSITE: PS00955; TCP1_3.
KW Chaperone; ATP-binding; 3D-structure; Complete proteome.
SQ SEQUENCE 543 AA: 58479 MW: F9DAEE63EB84E6 CRC64;
Query Match 11.58; Score 71.5; DB 1; Length 543;
Best Local Similarity 26.98; Pred. No. 20;
Matches 32; Conservative 15; Mismatches 59; Indels 13; Gaps 5;
DY 2 NNSLKEERVEEDNGKSD-----GNRGKPSDEV-VRTVTEEEVDEFEK-ILRRVHVHVAIR 53
b 339 SSDLGTAERVEQVKVEDYMTFTVTCGNPKAVSILVRGETEHVVDMSERTDLSLHVAS 398
DY 54 TVA-----KVGSGVAGELPSKKRKSQNLGRNSLDCNGVRGDEFEINRVGLQGLGLD 108
b 399 ALPDGAYAGGGATAAEZAFRLRSYAKIGRQQLAEIKFADA-IEEIPRALAENAGLD 456
RESULT 8
GREI_HUMAN
ID GREI_HUMAN STANDARD; PRT: 217 AA.
AC Q9HAV7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GrpE protein homolog 1, mitochondrial precursor (Mt-GrpE1) (HMG).
GN GREP1...
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE OF 2-117 FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=21211408; PubMed=11311562;
RA Choglay A.A., Chapple J.P., Blatch G.L., Cheetham M.E.;
RT "Identification and characterization of a human mitochondrial
RL homologue of the bacterial co-chaperone GrpE.";
RN Gene 267:125-134(2001).
RP SEQUENCE FROM N.A.
RL Li W.B., Graber C., Jessee J., Polayes D.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

```

-!- FUNCTION: COOPERATES WITH MITOCHONDRIAL HSP70 IN THE IMPORT
OF PROTEINS FROM THE CYTOLASM.
-!- SUBUNIT: BINDS TO HSP70, HSC70 AND HSP18.
-!- SUBCELLULAR LOCATION: Mitochondrial matrix.
-!- SIMILARITY: BELONGS TO THE GRPE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).

EMBL: AF298592; AAC31605.1;
EMBL: AL542571; -; NOT_ANNOTATED_CDS.
EMBL: BC024242; AAH24242.1;
HMM: P02248; 1AAR.
MIM: 606173;
InterPro: IPR000740; GrpE.
Pfam: PF01025; GrpE.
PRINTS: PR00773; GRPEPROTEIN.
PROSITE: PS01071; GRPE; 1.
Chaperone; Mitochondrial; Transit peptide.
TRANSIT 1 27 MITOCHONDRION (BY SIMILARITY).
CHAIN 28 217 GRPE PROTEIN HOMOLOG 1.
SEQUENCE 217 AA; 24279 MW; 40499AA3FC7E689E CRC64;

Query Match 11.4%; Score 71; DB 1; Length 217;
Best Local Similarity 25.6%; Pred. No. 8;
Matches 31; Conservative 16; Mismatches 32; Indels 42; Gaps 5;

y 9 ERVEEDNGSDGNCRKPSDEVYKIVTEBEV--DEFFKILRRVHVATRTVAKYNGVADCE 56
o : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
38 QNLBEDMGSGEOKADPPATE--KILLEEAKVLEQK-----ETVEKKRAIAIDIE 86
y 67 LPSSKKRRKSONGLRNS:PCNGVRHGFDEINRVLQGLGIDL-----NCKPEP 115
b : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
87 ---NLQRSSOKL-----VEEAKIYGIQAPKOLLEVDVLEKATQCVKPE 128
y 116 D 116
b 129 E 129

RESULT 9
C RA50_SULTO STANDARD; PRI: 879 AA.
D Q96Y85.
T 15-JUN-2002 (Rel. 41; Created)
T 15-JUN-2002 (Rel. 41; Last sequence update)
T 15-JUN-2002 (Rel. 41; Last annotation update)
E DNA double-strand break repair rad50 AtPase.
N RAD50 OR ST2108.
S Sulfolobus tokodaii.
C Archaea; Crenarchaeota; Thermoprotei; Sulfochaales; Sulfolobaceae;
C Sulfolobus.
N NCBI_TaxID=111955;
N [1]
P SEQUENCE FROM N.A.
C STRAIN-JCM 10345 / ?
X MEDLINE=21456156; PubMed=11572479;
A Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
A Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,
A Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Yamazaki M., Kato Y.,
A Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
A Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
A Oshima T., Kikuchi H.;
T *Complete genome sequence of an aerobic thermoacidophilic
T Crenarchaeon, Sulfolobus tokodaii strain7.*;
L DNA Res. 8:123-140(2001).
L -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
rad50/mre11 complex possesses single-strand endonuclease activity

```

```

CC      and 2/adult form; are produced by alternative splicing.
CC      -!- SIMILARITY: BELONGS TO THE TROPONIN T FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M10013; AAA49099.1;
CC      EMBL: K02263; AAA49098.1;
CC      PIR: A03086; TPCHTC.
CC      PIR: A25373; A25373.
CC      InterPro: IP001978; Troponin.
CC      Pfam: PF00992; Troponin; i.
CC      Muscle protein; Alternative splicing; Multigene family;
CC      Phosphorylation.
CC      INIT_MET 0 0 BY SIMILARITY.
CC      MOD_RES 1 1 PHOSPHORYLATION (BY CK2)
CC      VARSPIC 22 31 MISSING (IN ISOFORM 2).
CC      SEQUENCE 301 AA; 35854 MW; F85CEHA47F07DE94 CRC64;
CC      -----
CC      Query Match: 11.3%; Score 70.5; DB 1; Length 301;
CC      Best Local Similarity 28.7%; Pred. No. 13;
CC      Matches 33; Conservative 13; Mismatches 36; Indels 33; Gaps 6;
CC      -----
2Y 6 KXSERVEEDNGKSDGRGKSTWRTVEEVDKFKILRRVHVATRTVAKVNGS----51
b 22 EEELWEEGQGDQED-QVDEEERTFTTAEQDE-----TKANGCEGECEGR 67
y 62 ---VAEGSEKSKKRSKRSN:JGRNSLDCNGVRDGE---FDEINRVGLOGGLDLM 110
b 68 EQEPGECEKSKPKPKMPNLPVPPK-----LPDGERLDFDDTHRKWKE---KDLN 113
CC      -----
RESULT 11
IRI_VZVD
D RIRL_VZVD STANDARD: PRT: 775 AA.
C P09248;
T 01-MAR-1989 (Rel. 10, Last: sequence update)
T 01-MAR-1989 (Rel. 10, Last: sequence update)
T 01-JUN-1994 (Rel. 29, Last: annotation update)
E Ribonucleoside-diphosphate reductase large chain (EC 1.17.4.1)
E (Ribonucleotide reductase).
N 19.
S Varicella-zoster virus (strain Dumas) (VZV).
C Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Alphaherpesvirinae; Varicellovirus.
X NCB1_TaxID=10338;
N [1]
P SEQUENCE FROM N.A.
X MEDLINE=86306657; PubMed=3018124;
A Davison A.J., Scott J.E.;
T "The complete DNA sequence of varicella-zoster virus.";
J. Gen. Virol. 67:1759-1816(1986).
C -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
C -!- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized
C thioresoxin + H(2)O -> ribonucleoside diphosphate + reduced
C thioresoxin.
C -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
C -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
C -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE
C LARGE CHAIN FAMILY.
C -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----

```

```

CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: X04370; CAA27902.1;
CC      PIR: A27343; WMBE19.
CC      InterPro: IP000788; Ribonucleo_red.
CC      Pfam: PF00317; ribonuc_red_lg; 1.
CC      Pfam: PF02867; ribonuc_red_lgC; 1.
CC      PRINTS: PR01183; RIBORPTASEM1.
CC      PROSITE: PS00089; RIBORED_LARGE; 1.
CC      Oxidoreductase. DNA replication; Early protein.
CC      SEQUENCE 775 AA; 86827 MW; 181695A65E91B9D CRC64;
CC      -----
CC      Query Match: 11.3%; Score 70.5; DB 1; Length 775;
CC      Best Local Similarity 31.5%; Pred. No. 37;
CC      Matches 23; Conservative 7; Mismatches 14; Indels 29; Gaps 4;
CC      -----
2Y 51 AITVYAKVNGVGAEGELPSKKKRSKRSNGLRNSLDCNGVRDGEFDEINR---VGLQG---104
b 471 ARTVIFLNGVLAAGNPPCKK-----SKGVKN-----NRSGLGIGLGHHT 513
y 105 ----LGLDLNCKP 113
b 512 TCLRKGVDLTSP 524
CC      -----
RESULT 12
IF34_HUMAN
ID IF34_HUMAN STANDARD: PRT: 320 AA.
AC Q75921; O14801;
CT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last: sequence update)
DT 15-JUN-2002 (Rel. 41, Last: annotation update)
DE Eukaryotic translation initiation factor 3 subunit 4 (eIF3 delta)
DE (EIF3 p44) (eIF-3 RNA-binding subunit) (EIF3 p42).
GN EIF3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCB1_TaxID=9606;
XN [1]
XP SEQUENCE FROM N.A.
X MEDLINE=99841954; PubMed=9822659;
RA Block K.L., Vornlocher H.-P., Hershey J.W.B.;
RT "Characterization of cDNAs encoding the p44 and p35 subunits of human
RT translation initiation factor eIF3.";
RL C. Biol. Chem. 273:31901-31908(1998).
XN [2]
XP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
X MEDLINE=99141230; PubMed=9973622;
RA Bandhopadhyay A., Maitra U.;
RT "Cloning and characterization of the p42 subunit of mammalian
RT translation initiation factor 3 (eIF3): demonstration that eIF3
RT interacts with eIF5 in mammalian cells.";
RL Nucleic Acids Res. 27:1331-1337(1999).
CC -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA. THIS SUBUNIT BINDS TO THE 18S RRNA.
CC -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS.
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U96074; AAC78728.1;
CC      EMBL: AF020833; AAB71866.1;
CC      HSP: P19339; 2SXL.
CC      Genew: HNCN3274; EIF3A4.
CC      MIM: 603913.
CC      InterPro: IPR000504; RNA_rec_mot.
CC      -----

```

```
DR Pfam: PF00076; trm: 1.
DR SMART: SM00360; RRM: 1.
DR PROSITE: PS00102; RRM; 1.
DR PROSITE: PS00300; RRM_RNP_1; 1.
DR Initiation factor: Protein biosynthesis: RNA-binding.
DR DOMAIN 239 317 RNA-BINDING (RRM).
DR CONFLICT 293 293 K -> A (IN REF. 21).
DR CONFLICT 293 293 K -> A (IN REF. 21).
DR SEQUENCE 320 AA: 35696 MW: 70722626CECFB65FB CRC64;

Query Match 11.2% Score 70; DB 1; Length 320;
Best Local Similarity 26.1%; Pred. No. 15;
Matches 30; Conservative 17; Mismatches 26; Indels 42; Gaps 7;

QY 25 PSTEV-----VRTVTEVEDE---PFKILRRVIVATRTVAKVNGGVAEGFLPSKKHKRSCN 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 55 PKKEVINIKVTVEIKDECKKFKIVATRIETRKASK---AVA-----RKNKK 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 78 LQLRNSLDCNGVRGEF-----DEINRVGLQGLDQNKPRPESVS 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 105 FG-----NSEFDPGPNVATTVSDUVSMFTITSKF-DLNCQEEEDPMN 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
SP17_MOUSE
ID Q62252; STANDARD; PRT: 149 AA.
AC Q62252;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sperm surface protein Sp17 (Sperm autoantigenic protein 1?).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Taylor P., Page A.P., Kontopidis G., Husi H., Walkinshaw M.D.;
RA "The X-ray structure of a divergent cyclophilin from the nematode
RA parasite Brugia malayi.";
RA FEBS Lett. 425:361-366(1998).
RA [3]
RA X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS) OF 1-177.
RA MEDLINE=98218582; PubMed=9559680;
RA Taylor P., Page A.P., Kontopidis G., Husi H., Walkinshaw M.D.;
RA "The X-ray structure of a divergent cyclophilin from the nematode
RA parasite Brugia malayi.";
RA FEBS Lett. 425:361-366(1998).
RA [3]
RA X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 1-177.
RA S-RAIN-ATCC 75593;
RA MEDLINE=98318040; PubMed=9655334;
RA Mikol V., Ma D., Carlow C.K.S.;
RA "Crystal structure of the cyclophilin-like domain from the parasitic
RA nematode Brugia malayi.";
RA Protein Sci. 7:1310-1316(1998).
RA [4]
RA X-RAY CRYSTALLOGRAPHY (2.47 ANGSTROMS) OF 1-177.
RA MEDLINE=20108543; PubMed=10642184;
RA Ellis P.J., Carlow C.K.S., Ma D., Kuhn P.;
RA "Crystal structure of the complex of Brugia malayi cyclophilin and
RA cyclosporin A.";
RA Biochemistry 39:592-598(2000).
RA [1]
RA FUNCTION: PPIases accelerate the folding of proteins. It catalyzes
RA the cis-trans isomerization of proline imidic peptide bonds in
RA oligopeptides.
RA [1]
RA CATALYTIC ACTIVITY: Peptidylproline (omega-180) - peptidylproline
RA (omega=0).
RA [1]
RA ENZYME REGULATION: RELATIVELY INSENSITIVE TO INHIBITION BY CSA.
RA [1]
RA SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.
RA [1]
RA This SWISS-PROT entry is copyright. It is produced through a collaboration
RA between the Swiss Institute of Bioinformatics and the EMBL outstation
RA the European Bioinformatics Institute. There are no restrictions on its
RA use by non-profit institutions as long as its content is in no way
RA modified and this statement is not removed. Usage by and for commercial
RA entities requires a license agreement (See http://www.isb-sib.ch/announce/
RA or send an email to license@isb-sib.ch).
RA [1]
RA EMBL: 245299; CA886455.1;
RA HSSP: P12367; IR2A.
RA MGD: MGI:133378; Spal7.
RA InterPro: IPR000048; IQ region.
RA InterPro: IPR003117; RIIa.
RA Pfam: PF00612; IQ; 1.
RA Pfam: PF02197; RIIa; 1.
RA SMART: SM00615; IQ; 1.
RA SMART: SM00394; RIIa; 1.
RA PROSITE: PS50096; IQ; 1.
RA Membrane.
RA DOMAIN 112 141 IQ.
RA SEQUENCE 149 AA: 17296 MW: C7E05D11D6AFDC CRC64;
```


GenCore version 5.1.3
 Copyright (C) 1993 - 2002 CompuGen, Ltd.
 OM protein - protein search, using sw model
 Run on: December 4, 2002, 08:27:54 : Search time 82 seconds
 (without alignments)
 306,558 Million cell updates/sec

Title: US-09-733-685-2
 Perfect score: 623
 Sequence: 1 MNKSLKKEERVEEDNGKSDG.....QQLGELNCRPEPDSVLSL 122

Scoring table: BLOSUM62
 Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues
 Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database: SP-REMBL211

- 1: sp.archaea:
- 2: sp.bacteria:
- 3: sp.fungi:
- 4: sp.human:
- 5: sp.invertebrate:
- 6: sp.mammal:
- 7: sp.mhc:
- 8: sp.orcellule:
- 9: sp.phage:
- 10: sp.plant:
- 11: sp.priodent:
- 12: sp.virus:
- 13: sp.vertebrate:
- 14: sp.unclassified:
- 15: sp.virus:
- 16: sp.bacteriap:
- 17: sp.archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	623	100.0	122	10 Q9LJA3	Q9LJA3 Arabidopsis
2	96.5	15.5	85	10 Q82448	Q82448 Nicotiana t
3	80.5	12.9	265	16 Q9Z888	Q9Z888 chlamydia p
4	80.5	12.9	2254	10 Q9LNC2	Q9LNC2 Arabidopsis
5	80	12.8	753	2 Q9X553	Q9X553 neisseria m
6	79.5	12.8	2001	10 Q9M659	Q9M659 Arabidopsis
7	79.5	12.8	2001	10 Q9M658	Q9M658 Arabidopsis
8	78.5	12.6	265	16 Q9K2A4	Q9K2A4 chlamydia p
9	78	12.5	252	10 Q9PJF8	Q9PJF8 Arabidopsis
10	77.5	12.4	302	9 Q9TIE8	Q9TIE8 actinobacill
11	76.5	12.3	386	10 Q9FLD3	Q9FLD3 Arabidopsis
12	75.5	12.1	325	12 Q69272	Q69272 leporid her
13	75	12.0	384	2 Q9R3S1	Q9R3S1 actinobacill
14	75	12.0	365	2 Q9RAV2	Q9RAV2 actinobacill
15	75	12.0	375	2 Q9RCG2	Q9RCG2 actinobacill
16	75	12.0	404	5 Q95YR8	Q95YR8 leishmania

17	74.5	12.0	543	5 Q8SUE9	Q8SUE9 encephalito
18	74	11.9	141	10 Q48526	Q48526 Arabidopsis
19	74	11.9	307	17 Q26926	Q26926 methanobact
20	74	11.9	451	16 Q8XJ77	Q8XJ77 clostridium
21	74	11.9	891	10 Q22735	Q22735 Arabidopsis
22	73.5	11.8	389	10 Q9LQ22	Q9LQ22 Arabidopsis
23	73.5	11.8	611	5 Q9VSS1	Q9VSS1 drosophila
24	73.5	11.8	2045	16 Q9AOK5	Q9AOK5 streptococc
25	73	11.7	307	16 Q9JQ07	Q9JQ07 neisseria m
26	73	11.7	350	10 Q9L580	Q9L580 Arabidopsis
27	73	11.7	457	5 Q9Q222	Q9Q222 caenorhabdi
28	73	11.7	684	10 Q8RX78	Q8RX78 Arabidopsis
29	73	11.7	2149	10 Q9M3D3	Q9M3D3 Arabidopsis
30	72.5	11.6	464	17 Q981C9	Q981C9 sulfolobus
31	72.5	11.6	493	2 Q8VS92	Q8VS92 streptococc
32	72.5	11.6	873	5 Q9X556	Q9X556 caenorhabdi
33	72.5	11.6	919	5 Q9X555	Q9X555 caenorhabdi
34	72	11.6	132	3 Q9US24	Q9US24 schizosacch
35	71.5	11.5	167	5 Q9VYB0	Q9VYB0 drosophila
36	71.5	11.5	225	10 Q9AR85	Q9AR85 populus eur
37	71.5	11.5	249	5 Q962X5	Q962X5 drosophila
38	71.5	11.5	335	6 Q9HWV7	Q9HWV7 pseudomocas
39	71.5	11.5	551	5 Q95X34	Q95X34 caenorhabdi
40	71	11.4	467	10 Q9S9R6	Q9S9R6 Arabidopsis
41	71	11.4	707	10 Q9AUS0	Q9AUS0 oryza sativ
42	71	11.4	783	12 Q9E1Z6	Q9E1Z6 cercopithe
43	71	11.4	862	10 Q80468	Q80468 Arabidopsis
44	71	11.4	1859	10 Q94H07	Q94H07 oryza sativ
45	70.5	11.3	314	5 Q21760	Q21760 caenorhabdi

ALIGNMENTS

RESULT 1:
 Q9LJA3
 ID Q9LJA3 PRELIMINARY: PRT: 122 AA.
 AC Q9LJA3
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 15, Last annotation update)
 DE Genomic DNA, chromosome 3, P1 clone: MPE11 (NIMIN-2 protein).
 GN NIMIN-2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 CX NCBI_TaxID=3702;
 FN [1]
 PP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Sato S., Nakamura Y., Kaneko T., Kato I., Asamizu E., Tabata S.,
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 PP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty P1 and PAC
 RT clones.";
 RL DNA Res. 7:131-135(2000).
 RN [3]
 PP SEQUENCE FROM N.A.
 RC STRAIN=CV. COL-0; TISSUE=GREEN VEGETATIVE TISSUE;
 RA Weigel R.R., Pfitzner A.J.P., Pfitzner U.M.;
 RT "NIMIN-1, NIMIN-2 and NIMIN-3, members of a novel family of proteins
 RT from Arabidopsis that interact in vivo with NPR1/NIM1, a key regulator
 RT of systemic acquired resistance in plants.";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB023041; BAB01050.1; ...
 DR EMBL: AJ220185; CAC19845.1; ...
 SO SEQUENCE 122 AA; 13508 MW; 24FD70920CE1C152 CRC64;


```
DR HSP; P00918; Ibv3.
OR InterPro; IPR001143; Euk_COanhd.
OR Pfam; PF00194; Carb_anhydrtase; 1.
OR ProDom; PD000865; Euk_COanhd; 1.
SQ SEQUENCE 2254 AA; 248158 MW; 64E80CE274731FF3 CRC64;

Query Match 12.9%; Score 80.5; DB 10; Length 2254;
Best Local Similarity 21.7%; Pred. No. 85;
Matches 25; Conservative 25; Mismatches 34; Indels 31; Gaps 4;

QY 4 LUKKEERV-----EEDNGKSDGNRGKPSCTEVVKTIVIEEVDFF 42
Db 232 NMKKDKIGLTGRTIYTRSLAASIPASVEQETPGLRRSSRGTPSTKVIITPASATRKSE-- 289
QY 43 KILRRVHVATRTVAKVNGVAGELPSKPKRSQ---NLGLRNSLDC-NGVRDGE 93
Db 290 ----RLAPSPASVSKSGGIVKNSTPSSLRRNRGKTEVSLOSSKSGSDNSIRKGD 340

RESULT 5
Q9X5B3 PRELIMINARY: PRT: 753 AA.
ID Q9X5B3
AC Q9X5B3;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Lactoferrin-binding protein.
SN LBPB.
SC Neisseria meningitidis.
OS Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=487;
DX (1)
RX SEQUENCE FROM N.A.
RP STRAIN-M990;
RC MEDLINE=99250255; PubMed=12231574;
RA Pettersson A., van der Brien C., Joosten V., Hendriksen J.,
RA Tomassen J.;
RT "Sequence variability of the meningococcal lactoferrin-binding protein
RT LbpB";
RL Gene 231:105-110(1999).
OR ENBL; AF123381; A0431769.1; -.
OR InterPro; IPR001677; Transferrin_bind.
OR Pfam; PF01298; Lipoprotein_5; 1.
Q SEQUENCE 753 AA; 82180 MW; CE2670AB0389AALC CRC64;

Query Match 12.8%; Score 80; DB 2; Length 753;
Best Local Similarity 27.8%; Pred. No. 25;
Matches 25; Conservative 18; Mismatches 33; Indels 14; Gaps 3;

Y 6 KKEERVEEDNGKSDGNRGKPSCTEVVKTIVIEEVDFFKILRRVHVATRTVAKVNGVAGEG 65
b 496 EEDTSEEDNGEDEATAEETEVEDEAEVEE-----PEKSPAEENGCG-SGS 545
Y 66 ELPSKKRKSQNLGLRNSLDCNGVRDGEFD 95
b 546 ILPALEASKSGROI-----DPLKIGRTAETD 571

RESULT 6
Q9M659 PRELIMINARY: PRT: 2001 AA.
ID Q9M659;
AC Q9M659;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MOM.
SC Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C eucosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
DX (1)
RX SEQUENCE FROM N.A.
RP STRAIN-M990;
RC MEDLINE=20792299; PubMed=10821279;
RA Amedeo P., Habu Y., Afsar K., Mittelsten Scheid O., Paszkowski J.;
RA "Disruption of the plant gene MOM releases transcriptional silencing
RT of methylated genes";
RC Nature 405:203-206(2000).
DR ENBL; AF213627; AAF73380.1; -.
Q SEQUENCE 2001 AA; 218586 MW; F6C5DAE897F2FE61 CRC64;

Query Match 12.8%; Score 79.5; DB 10; Length 2001;
Best Local Similarity 21.9%; Pred. No. 92;
Matches 25; Conservative 24; Mismatches 34; Indels 31; Gaps 4;

QY 5 LKKEERV-----EEDNGKSDGNRGKPSCTEVVKTIVIEEVDFFK 43
Db 1 MKKDEKIGLTGRTIYTRSLAASIPASVEQETPGLRRSSRGTPSTKVIITPASATRKSE--- 57
QY 44 ILRRVHVATRTVAKVNGVAGELPSKPKRSQ---NLGLRNSLDC-NGVRDGE 93
Db 58 ----RLAPSPASVSKSGGIVKNSTPSSLRRNRGKTEVSLOSSKSGSDNSIRKGD 108

RESULT 7
Q9M658 PRELIMINARY: PRT: 2001 AA.
ID Q9M658;
AC Q9M658;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MOM.
SC Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C eucosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
DX (1)
RX SEQUENCE FROM N.A.
RP STRAIN-M990;
RC MEDLINE=20792299; PubMed=10821279;
RA Amedeo P., Habu Y., Afsar K., Mittelsten Scheid O., Paszkowski J.;
RA "Disruption of the plant gene MOM releases transcriptional silencing
RT of methylated genes";
RC Nature 405:203-206(2000).
DR ENBL; AF213628; AAF73381.1; -.
Q SEQUENCE 2001 AA; 218569 MW; B9C85D4E62704441 CRC64;

Query Match 12.8%; Score 79.5; DB 10; Length 2001;
Best Local Similarity 21.9%; Pred. No. 92;
Matches 25; Conservative 24; Mismatches 34; Indels 31; Gaps 4;

QY 5 LKKEERV-----EEDNGKSDGNRGKPSCTEVVKTIVIEEVDFFK 43
Db 1 MKKDEKIGLTGRTIYTRSLAASIPASVEQETPGLRRSSRGTPSTKVIITPASATRKSE--- 57
QY 44 ILRRVHVATRTVAKVNGVAGELPSKPKRSQ---NLGLRNSLDC-NGVRDGE 93
Db 58 ----RLAPSPASVSKSGGIVKNSTPSSLRRNRGKTEVSLOSSKSGSDNSIRKGD 108

RESULT 8
Q9K2A4 PRELIMINARY: PRT: 265 AA.
ID Q9K2A4;
AC Q9K2A4;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 18, Last annotation update)
DE Hypothetical protein CP0291.
SN CP0291.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83558;
DX (1)
RX SEQUENCE FROM N.A.
RP STRAIN-M990;
RC MEDLINE=20792299; PubMed=10821279;
RA Amedeo P., Habu Y., Afsar K., Mittelsten Scheid O., Paszkowski J.;
RA "Disruption of the plant gene MOM releases transcriptional silencing
RT of methylated genes";
RC Nature 405:203-206(2000).
DR ENBL; AF213629; AAF73382.1; -.
Q SEQUENCE 2001 AA; 218569 MW; B9C85D4E62704441 CRC64;
```

```
RC STRAIN-AR39;
RX MEDLINE-20150255; PubMed-10684935;
RA Read I.D., Brunham R.C., Sten C., Gill S.R., Heidelberg J.F.,
RR White O., Hickey E.K., Peterson J., Otterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RR Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AE002191; AAF38248.1;
OR TIGR: CP0291;
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 25523 MW; FA9B4AFEBCEFC30FE CRC64;

Query Match 12.6%; Score 78.5; DB 16; Length 265;
Best Local Similarity 27.4%; Pred. No. 10;
Matches 40; Conservative 18; Mismatches 37; Indels 51; Gaps 7;

CY 13 EDNGKSGNRCK-----PSTEVRTVTEEE---VDEFFKTLR-RVHVATRTVAKVSGV 52
DB 105 EDLGVKGRTPEDPVITPTPEIVOLIPDEELSTVDEALQGRSLRYAVKSVK----- 159
CY 63 ABEGEPKRRKRSQGLRNSL-----CNGVR-----DGEFD 95
DB 160 -----PMQDLALVGFGLRDSAGLNFVRLANGVQNHYPHTKVKLYLAKNLAQWVDEIS 214
CY 96 EINRVGLQGLGDLNCKPEPVSUS 121
DB 215 EEKQRLALGLD-----PKIESISIT 236

RESULT 9
19-JF8 PRELIMINARY: PRT: 252 AA.
AC Q9FJF8;
T 01-MAR-2001 (TrEMBLrel. 16, Created)
T 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
T 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
E Genomic DNA, chromosome 5, pl clone:MCM23 (Hypothetical 28.5 kDa protein).
S Arabidopsis thaliana (Mouse-ear cross).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C eurosids II; Brassicales; Brassicaceae; Arabidopsis.
X X NCBI_TaxID=3702;
N [1]
P SEQUENCE FROM N.A.
C STRAIN-COLUMBIA;
X MEDLINE-99087469; PubMed-9872454;
X Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
X Tabata S.;
A "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
T Sequence features of the regions of 1,013,767 bp covered by sixteen
T physically assigned P1 and TAC clones.";
L DNA Res. 5:297-308(1998).
P [2]
P SEQUENCE FROM N.A.
A Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Gale J.M.,
A Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Qiach H.L.,
A Tang C., Toriumi M., Yandura Y., Yu G., Yu S., Bowser J.,
A Kamiya P., Chen H., Cheuk R., Hayashizaki Y., Ishida S., Jones T.,
A Kambali A., Karlin-Neumann G., Kawai J., Kim C., Koesoma E., Lue R.,
A Lit J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
A Sakurai T., Satou M., Seki M., Shino M., Southwick A., Tracy S.E.,
A Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
T "Full length cDNA of gene MCM23.14/AF15916040 (GI:9757896).";
L Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
L EMBL: AB015473; BAB98403.1;
R EMBL: AY050965; AAK93642.1;
W Hypothetical protein.
Q SEQUENCE 252 AA; 28549 MW; 78DF092FA6C60D7F CRC64;
```

```
Query Match 12.5%; Score 78; DB 10; Length 252;
Best Local Similarity 39.3%; Pred. No. 10;
Matches 24; Conservative 10; Mismatches 17; Indels 10; Gaps 3;

CY 26 EVVRTVTEEVDEFFKIL-----RVHVATRTVAKVNGVAGVGLPSKKRRKRSQNLGLR 81
DB 131 EVIEF---KSLDEAFKLLIKQPGARLHVFSFDLQNGVEGVYEG-LSSLRKRESRYVGLR 186
CY 52 N 62
DB 187 D 187

RESULT 10
Q9TIF8 PRELIMINARY: PRT: 302 AA.
AC Q9TIF8;
T 01-MAY-2000 (TrEMBLrel. 13, Created)
T 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
T 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
D Hypothetical 32.9 kDa protein.
O Lactobacillus bacteriophage phi adh.
OC Lactobacillus bacteriophage phi adh.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=12417;
RN [1]
RP SEQUENCE FROM N.A.
RA Altermann E.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-99384014; PubMed=10452953;
RA Altermann E., Klein J., Heinrich B.;
RT "Primary structure and features of the genome of the Lactobacillus
RT gasserii temperate bacteriophage phi-adh.";
RI Gene 236:333-346(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-95138034; PubMed=7836307;
RA Heinrich B., Binshofer B., Blaesi J.;
RT "Primary structure and functional analysis of the lysis genes of
RT Lactobacillus gasserii bacteriophage phi-adh.";
RL J. Bacteriol. 177:723-732(1995).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE-93231538; PubMed=8472961;
RA Tremaux C., de Antoni G., Raya R., Klaenhammer T.;
RT "Genetic organization and sequence of the region encoding integrative
RT functions from Lactobacillus gasserii temperate bacteriophage phi-
RT adh.";
RL Gene 126:61-66(1993).
RN [5]
RP SEQUENCE FROM N.A.
RA Kage G., Altermann E., Klein J., Heinrich B.;
RT "Structure of a genome region of the Lactobacillus gasserii temperate
RT phage phi adh covering a repressor gene and cognate promoters.";
RL Gene 210:67-70(1998).
DR EMBL: AJ131519; CAB52529.1;
KW Hypothetical protein.
SQ SEQUENCE 302 AA; 32916 MW; 6C953088D1A83109 CRC64;

Query Match 12.4%; Score 77.5; DB 9; Length 302;
Best Local Similarity 24.8%; Pred. No. 15;
Matches 32; Conservative 26; Mismatches 38; Indels 33; Gaps 5;

CY 1 MNNSLAKKEKVEEDNCKSGNRKGPSTEVRTVTEEVDEFFKILRRVHVATRTVAKVNG 50
DB 63 LGUTLKKQOSSLERN-KSELSLKEAQSKVDKSTKGRNEYERYSKYKATAERNVANATT 121
CY 61 GVAEGELPSKKRRKRSQNLRLNSLDC--NGV-----RDGEFDE 96
DB 122 RIAK---LSCQQDRA-----RNSLDYKSGLASAQSALKKISESSNAVYGRLEAEKHEE 173
CY 97 INEVGIQGL 105
```


Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Actinobacillus.
NCBI_TaxID=715;
[1]
SEQUENCE FROM N.A.
STRAIN-WF83;; Pubmed-9809431;
MEDLINE-98026959; PMID-9809431;
ITO H., Osaki M., Tchiida I., Ohya T., Sekizaki T.;
"Demonstration of the third antigenically distinct outer membrane
protein (OmpA) in Actinobacillus pleuropneumoniae serotype 7";
JEMS Microbiol. Lett. 157:303-308(1998).
[2]
SEQUENCE FROM N.A.
STRAIN-WF83;
MEDLINE-9812623;; PubMed-9466755;
Gram T., Ahrens P.;
"Improved diagnostic PCR assay for Actinobacillus pleuropneumoniae
based on the nucleotide sequence of an outer membrane lipoprotein";
J. Clin. Microbiol. 36:443-448(1998);
EMBL: AB007579; BAA34662.1;
EMBL: U86682; AAC00615.1;
InterPro: IPR005014; Lipoprotein_14;
Pfam: PF03346; Lipoprotein_14; 1.
Lipoprotein; Signal.
SIGNAL 1 19 POTENTIAL.
SEQUENCE 365 AA; 49708 MW; 30FA543CC115C9F3 CRC64;

[illegible]

```

RESULT 15
          5RCG2
D      PRELIMINARY;          PRT:   375 AA.
C      Q9RCG2;
T      01-MAY-2000 (TREMUREL_13, Created)
I      01-MAY-2000 (TREMUREL_13, Last sequence update);
I      01-JUN-2002 (TREMUREL_21, Last annotation update)
E      Outer membrane lipoprotein.
S      Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
C      Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.
X      Actinobacillus.
X      NCBI_TaxID=715;
X      [1]
N      SEQUENCE FROM N.A.
P      STRAIN=M62;
A      MEDLINE=98126231; PubMed=9466755;
X      Gram T., Ahrens P.;
X      *Improved diagnostic PCR assay for Actinobacillus pleuropneumoniae.
I      based on the nucleotide sequence of an outer membrane lipoprotein.
L      J. Clin. Microbiol. 36:443-448(1998).
R      EMBL: U96678; AAD00611.1;
R      InterPro: IPR005014; Lipoprotein_14.

```

```

DR  PRAM: PF03346; Lipoprotein_14; 1.
SQ  Lipoprotein_14.
KW  SEQUENCE 375 AA: 40840 KW: 67BA62BE56E01A02 CRC64:
      Query Match 12.0%; Score 75; DB 2: Length 375;
      Best Local Similarity 24.6%; Pred. No. 33;
      Matches 31; Conservative 21; Mismatches 56; Indels 16; Gaps
QY  6  KSEVEREENKSGDGNRGK-----PSTEVRRVTVTEEEVDFFKLRVHVHAKR 53
DB  120  KDDQKLEEPNKNSEAEILKELGIKQIKTGIITRSDVVLNLTDE-QENIQRLSESDIVR 178
QY  54  TVAKVGVGVAEGELPSKKKRKRNGLNRLSDCNGVRDGEFDEINRVGLQG---LGLLN 110
DB  179  NDLKIITMIPNQDIRTLKSDTGLLCGYCYGMQLNQVREGERYGINVDLVGHYLLSMUES 238
QY  111  CKPEPD 116
DB  239  TKTAPN 244
      Search completed: December 4, 2002, 09:25:04
      Job time : 94 secs

```

Search completed: December 4, 2002, 09:25:04
Job time : 94 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Inc.

DN nucleic - nucleic search, using sw model

un on: December 4, 2002, 07:05:01 : Search time 3076 seconds
(without alignments)
16094.125 Million cell updates/sec

title: US-09-733-685-3

effect score: 1700

sequence: 1 tgggtttttattgataaca.....tttacgaacttaacaacatac 1700

coring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

searched: 2054640 seqs, 14551402878 residues

total number of hits satisfying chosen parameters: 4109280

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database :

GenEmbl:

1: gb_ba:

2: gb_ba:

3: gb_ba:

4: gb_ba:

5: gb_ba:

6: gb_ba:

7: gb_ba:

8: gb_ba:

9: gb_ba:

10: gb_ba:

11: gb_ba:

12: gb_ba:

13: gb_ba:

14: gb_ba:

15: gb_ba:

16: gb_ba:

17: gb_ba:

18: gb_ba:

19: gb_ba:

20: gb_ba:

21: gb_ba:

22: gb_ba:

23: gb_ba:

24: gb_ba:

25: gb_ba:

26: gb_ba:

27: gb_ba:

28: gb_ba:

29: gb_ba:

30: gb_ba:

31: gb_ba:

32: gb_ba:

33: gb_ba:

34: gb_ba:

35: gb_ba:

36: gb_ba:

37: gb_ba:

38: gb_ba:

39: gb_ba:

40: gb_ba:

41: gb_ba:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1700	100.0	83650	8	AB023041 Arabidops
C 2	576.2	33.9	92620	8	AB026636 Arabidops
C 3	540.2	31.8	83646	8	AB005248 Arabidops
C 4	534.6	31.4	94487	8	AC012394 Arabidops
C 5	534.6	31.4	100806	8	AC015450 Arabidops
C 6	499.8	29.4	104386	8	AL161813 Arabidops
C 7	499.8	29.4	179771	8	AT132A17 Arabidops
C 8	497	29.2	95519	8	AF071527 Arabidops
C 9	497	29.2	116448	8	AC005142 Arabidops
C 10	497	29.2	159629	8	ATCHR1V9 Arabidops
C 11	369	21.7	369	8	ATR250185 Arabidops
C 12	330.8	19.5	95190	8	AC007203 Arabidops
C 13	103.2	6.1	7218	6	166494 Sequence 14
C 14	90.2	5.3	171763	2	AC115878 Mus muscu
C 15	89.2	5.2	280982	2	AC125184 Mus muscu
C 16	87.4	5.1	163362	2	AC101939 Mus muscu
C 17	87.4	5.1	217779	2	AL844515 Mus muscu
C 18	86.8	5.1	202872	2	AC016160 Homo sapi
C 19	85.6	5.1	166299	2	AC127244 Mus muscu
C 20	85.6	5.0	44735	9	AC092315 Homo sapi
C 21	85.6	5.0	48623	2	AC130469 Homo sapi
C 22	85	5.0	178036	10	AL672308 Mouse DNA
C 23	84.6	5.0	422	5	PD0M04 X93505 P. domestica
C 24	84.6	5.0	51953	2	AC101966 Mus muscu
C 25	84.6	5.0	177648	10	AL732328 Mouse DNA
C 26	84.6	5.0	228944	2	AL844530 Mus muscu
C 27	84.4	5.0	222556	2	AC126558 Mus muscu
C 28	84	4.9	417	5	AF044763 Cecropia
C 29	84	4.9	202756	2	AC119804 Mus muscu
C 30	84	4.9	218898	2	AC114903 Mus muscu
C 31	83.6	4.9	162803	2	AC117818 Mus muscu
C 32	83.6	4.9	265538	2	AC102446 Mus muscu
C 33	83.4	4.9	239851	2	AC101712 Mus muscu
C 34	83.4	4.9	247340	2	AC121591 Mus muscu
C 35	83.2	4.9	64789	2	AC083839 Homo sapi
C 36	82.8	4.9	167364	10	AC121590 Mus muscu
C 37	82.8	4.9	220275	10	AL732404 Mouse DNA
C 38	82.6	4.9	48699	2	AC115677 Dictyoste
C 39	82.4	4.8	178413	2	AL805954 Mus muscu
C 40	82.4	4.8	203295	10	AL663049 Mouse DNA
C 41	82.2	4.8	204829	2	AC113318 Mus muscu
C 42	82	4.8	163132	2	AC097180 Rattus no
C 43	81.6	4.8	93661	9	AC107463 Homo sapi
C 44	81.6	4.8	106815	10	AL645809 Mouse DNA
C 45	81.6	4.8	120635	2	AC127463 Ornithorn

ALIGNMENTS

RESULT 1
LOCUS AB023041/c 83650 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, Pl clone: MPE11.
ACCESSION AB023041 BA000014
VERSION AB023041.1 GI:4220640
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui Pl clone:MPE11.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (sites)
AUTHORS Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.

Pred. No. is the number of results predicted by chance to have a

FILE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
JOURNAL
COMMENT

Structural analysis of Arabidopsis thaliana chromosome 3. 1.
 Sequence features of the regions of 4,504,864 bp covered by sixty
 P1 and TAC clones
 DNA Res. 7 (2), 131-135 (2000)
 2 (bases 1 to 83650)
 Sato, S., Nakamura, Y., Kaneke, T., Kato, T., Asamizu, F. and Tabata, S.
 Direct Submission:
 Submitted (01-FEB-1999); Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research: 1532-3, Yana,
 Kisarazu, Chiba 252-0612, Japan (E-mail: ynakamura@kazusa.or.jp,
 Tel: 81-438-52-3335, Fax: 81-438-52-3334)
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/sgd-graph.cgi?c=MPE11
 Genes with similarity to proteins in the databases are described as
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Graal
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/Graal-1.3/),
 GENSCAN (Chris Burge, MIT, http://CCR-061.mit.edu/GENSCAN.html),
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://grembl.zool.tastata.edu/cgi-bin/sp.cgi)
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
 http://genome.wustl.edu/eddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is K9122 and the 3' clone is M2114.
 Location/Qualifiers
 1..83650
 /organism="Arabidopsis thaliana"
 /strain="Col-0mbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="MPE11"
 /clone_lib="Mitsui pl"
 complement(542..764)
 /note="CDS is reported in Acc# AP000599
 contains similarity to CHP-rich zinc finger protein
 gene_id:K9122.5"
 /number=1
 /evidence=not_experimental
 join(1765..2463,2548..2739,2662..2649,3037..3107,
 3190..3273,3410..3514)
 /note="gb:AA055139.1
 gene_id:MPE11.1"
 /codon_start=1
 /evidence=not_experimental
 /product="dihydrolipomate S-acetyltransferase"
 /protein_id="BAB01047.1"
 /db_xref="GI:9279589"
 /translation="MTVSKTRTEIFPALSSTMIETGKIVSKIKTESKIAKGSVVVV
 ESDKADMDVFEDYSGLAAIVVGETAPVGAATGILAEPAIEEAKSKRAKSSSS
 VAERAVPSPPVTSAPATAPVAVSGCAKTVATPKAKAKQKAVGIESVAG
 TGPGRITASDVERAAGIAKSKSLAPPPPPPTAKATINPLPUSSIVPTA
 MQSAVKNMIESLVPTFRVGPVNTDALDLYEKVKPKGVMTA:LAKAKAGMALAJH
 PVNAACKDGKGSINIAVAINGSLITPIVQADKLDLYLSSQKKELVGKA
 RSKLOPHVYNSGTFTLSNLGMFGVDRFDAILPPGAGINAVAGSKPTVADKGFPS
 VKNTMLVNTAHRIVVCAADLAALFIQTFAKLIENPDSITL"
 complement(4594..5106)
 /note="gene_id:MPE11.2
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01048.1"
 /db_xref="GI:9279590"
 /translation="MEGLLEELTPTDSVGNKRVKRDGLDLSPOVKRLEKLLFDSDSL
 DPVSODLDSVKNKSFENSLTAAISSGTUPLGGLFRASDDGLGPPPTPTCTLI
 PPSCEETVELVRASSDSSEYGE:CGEDHVTETFCPLGSDGGLFEYFDGCLSDGLF

FEATURES
source
exon:
CDS
CDS

SRRPELPAE"
 complement(join(10090..10408,10649..10804,10916..10962))
 /note="gene_id:MPE11.3
 similar to unknown protein
 sp|P42744"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01049.1"
 /db_xref="GI:9279591"
 /translation="WNEPKAYDRLQMTYIQTLEASICLLNCGPIGSKNKLNLVLS
 VGVSITIVGSKVLIGDKWQFHRHAEQKFSISEGFRDENTYFQRSEQSVFQRQ,
 EGNRIAGO:VRPMQIARRIWAISIGKWSLADNTITYGKEARAITDPPGRVLRLTS:
 VDRHLDPSFLCDI"
 complement(11450..11818)
 /note="gene_id:MPE11.4
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01050.1"
 /db_xref="GI:9279592"
 /translation="MNSLKEERVEEDNGSKDGNRKPSTEVVTVTEVEEDVEFKI
 LRHVAVRTVAKVNGVAEGELPSKRRKRSNGLRNSLDCNGVRDGEFDEINRVLG
 QGLGLDIACKPEDPSVLSL"
 1698..17979
 /note="gene_id:MPE11.5"
 /codon_start=1
 /evidence=not_experimental
 /product="AP2 domain transcription factor-like protein"
 /protein_id="BAB01051.1"
 /db_xref="GI:9279593"
 /translation="MAERKKRSIOTNPKPKMKKPPQLNHLPLGLSDLTKMKLR
 FVNDPYATYSSSEERSORRYVCEILDLPFAQAATOAESSESSYCOESNNNGSK
 TKISCSKVLRSKASPVVGRSTTVSPVGVQRKWKAAEIRHPITKRVILGTY
 EILEGAADATKLEFDALAAATSAASVLSNESMSISGSSIDJCKLVDSTLD
 QQAGSKASPDFADLQIPMGCFIDDFIPNACELDFLTENNQMDDYCGID
 DLDI:GLECDGSPDVFSDVLDLGIITIDKYAFVDHIATITPTLNIACP"
 join(21893..22063,22330..22386,22578..22783,22894..23227,
 23311..23382,23473..23613)
 /note="gb:AAAF23821.1
 gene_id:MPE11.6"
 /codon_start=1
 /evidence=not_experimental
 /product="homocysteine S-methyltransferase AtHM-1"
 /protein_id="BAB01052.1"
 /db_xref="GI:9279594"
 /translation="MVLEKSKALLEDLTKKCGCAVVGGFATQLEIGHAALNDPLKS
 AVS:KNPELIRVHMYLEAGADIVTSSYQATIPGLSRGLSEESLQKSVEL
 AVEARDREKVKVSGSHSYNRYALVAASIGSYGAYLADSEYSHYGENVSLQKDF
 HRRRLQVLEAGPOLLAFTIPNLEAQACVELLEEKVQIPAWICFTSVGEEKAPSG
 ESFECLPLKNSNIYAVGINCAPQFTENIRKPAKTKKAIIVVYNSGEVWDGKA
 KQWLPSCFGDDPEFMFATKWRDLGAKLLIGGCCRTTPTSTINAIKRLKRR"
 complement(27277..27552)
 /note="gene_id:MPE11.7
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01053.1"
 /db_xref="GI:9279595"
 /translation="MTHAREKSSLTITLLWLLSYMLHFCVYSRYCAIMHPTETP
 ASGRKEEDLMKRYFOAGKEFPVDFGVGSGESKRIVPSCDPLRN"
 30765..31883
 /note="gb:AAAF16548.1
 gene_id:MPE11.8
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01054.1"
 /db_xref="GI:9279596"
 /translation="MPKEREKRSVLDKYKRSPLCCEASLAKPSEKQVKEWAEARC2
 VMEH:PHNIGILLICSSYNGCRPYMCDTSHRSNCFDFPKASKPKPSLLRBEES
 NPTMEDVDSSTAVNLGEAASEITVVDLSDGERGEVEEVEEVEEVEEVEEVEE
 ITTEDEKKNPKQKLCPLCRGHIKENYVKAARCFMNSKRSCTCETCDFSGSYDLIR
 KHARLLHPCVRPSEADPQRQRWRRLERQSLGCLLSTLQSSFGGDEISNDGDFLAD

CDs
IRLLIVYLIRVFPESGSSSSWSGTSRAKTHISGRKSSRSPASLWSESYGENTUT
SPROENQSSDEQVSGTTRRRSRRTVVDDEDEEF
complement(join(32197..32430,32525..32676,32969..33332,
33464..33487))
/note="gene_id:MPell.9"

/codon_start=1
/evidence-not_experimental
/product="chloroplast 50S ribosomal protein L15"
/protein_id="BAB01055.1"
/db_xref="GI:9279597"
/translation="MAPLSLSSNPLSRHCYRLHLSSPKGNVSVLGNPSSQILSL
KLNQIKTRNQOFARE:VYVSQTAATSSAVYAPERLNLDPQPSRKKQKRGK
ISAGASCGMGKSGPGLMRGEGGTALYRLRLKRLGIAGMSRGLPKLP
VNKIDTAGDEGDSLEKTLKQGLNPSGRERKLPKLTIGIGELSMKLTFAKAP
STOAKRELEAGSGLTLVLPGRKKVVPKSVAKNQARADEYFAKRRAAAEAAATSEPAS
A"

CDs
complement(join(33961..34021,34144..34226,34391..34482,
34602..34800))
/note="db:BAF26483.1
gene_id:MPell.10
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01056.1"
/db_xref="GI:9279598"

CDs
/translation="MKNVMLIDESNASYDLINWLENOKDITSSSVYFAKQPNS
ETPPVSSVGFQIYFPSPNSQLRFLKGRKNTALGLEKAKICLNHG:KAE:
FDDGPKDLJRKIQEINILNIVTSQOOLKQNTQDCSLVVKRQRKD"
join(36033..36259,36243..36283,36497..36565,36651..36707,
36797..36871)

CDs
/note="contains similarity to transcription-associated
zinc ribbon protein
gene_id:MPell.11"

/codon_start=1
/evidence-not_experimental
/protein_id="BAB01057.1"
/db_xref="GI:9279599"
/translation="MEKRSSEFLFCNLGMLVLSKTYAECPRKTTNKKILUK
SIATVSAEDIRRELGLSLFCEKTOAEALPKIKKACEKQHPVLVTTTC:TSANEG
QTYTTCPCAHRETEG"
join(37285..37965,38579..38674)

CDs
Query Match 100.0% Score 1700; DB 8; Length 83650;
Best Local Similarity 100.0%; Pred. No. 2.1e-285;
Matches 1700; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 TGGGTTTATGTAACAGACAAATATTATTATTCATGAGTTTATTGATAG 60
DB 12680 TGGGTTTTATTGCAATACATGACAAATATTATTATTTCATGAGTTTATTGATAG 12621
Y 61 CATGACAAATATTAATA:CAAGTGTAAATAACATGTTTGTCTTAAATAACATGCAT 120
DB 12620 CATGACAAATATTAATA:CAAGTGTAAATAACATGTTTGTCTTAAATAACATGCAT 12561
Y 121 TTAANATCAGACATTTGTTTAAATGAATCAATCTTATATATATCAACAGACATTGAC 190
DB 12560 TTAANATCAGACATTTGTTTAAATGAATCAATCTTATATCAACAGACATTGAC 12501
Y 181 GGAATAATTCAGGTAAAGAGAGAAATAAGAAATAGAGATAGAGAGATTCTATGGAATA 240
DB 12500 GGAATAATTCAGGTAAAGAGAGAAATAAGAAATAGAGATAGAGAGATTCTATGGAATA 12441
Y 241 AGAAGAGAGACATGTAGGTGAACAAAATAAAGAGATATGATGATAATTTATGAGAG 300
DB 12440 AGAAGAGAGACATGTAGGTGAACAAAATAAAGAGATATGATGATAATTTATGAGAG 12381
Y 301 GTGCTGAAGATTATTTAGAGAGGAGAGAGAAATAGAAAGAAATGACATGCTGGA 360
DB 12380 GTGCTGAAGATTATTTAGAGAGGAGAGAGAAATAGAAAGAAATGACATGCTGGA 12321
Y 361 TCTGAAGAGATGAATTTGCTTAAAGATGAAGAGAGAAAGAGAAATCCATGCTTAAAGTC 420
DB 12320 TCTGAAGAGATGAATTTGCTTAAAGATGAAGAGAGAAAGAGAAATCCATGCTTAAAGTC 12261

```
1501 GGCCTTTGGCTGGCCGCTTACTATAATTGTCAGAGTAGCTAIGGCCTGTATAAGGAGT 1560
|||||
11180 GGCCTTTGGCTGGCCGCTTACTATAATTGTCAGAGTAGCTAIGGCCTGTATAAGGAGT 11221
|||||
1561 CCAAAAAGAGCGGCGCAATTCGCGTTCGGTTCGCTTTGGAACCTTGGATTGCGATTAGT 1620
|||||
11120 CCAAAAAGAGCGGCGCAATTCGCGTTCGGTTCGCTTTGGAACCTTGGATTGCGATTAGT 11061
|||||
1621 CATGTTTATCATTAATGCTCGGACCTTGGACGACGGCTTGTTCTTCCTCTG 1680
|||||
11060 CATGTTTATCATTAATGCTCGGACCTTGGACGACGGCTTGTTCTTCCTCTG 11001
|||||
1681 TTACGACTACGACATAT 1700
|||||
11000 TTACGACTACGACATAT 10981
|||||

RESULT 2
LOCUS AB026636 92620 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone:K14A17.
ACCESSION AB026636 BA000014
VERSION AB026636.1 GI:4757392
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui TAC clone:K14A17.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 3. I.
Sequence features of the regions of 4,504,864 bp covered by sixty
P1 and TAC clones
DNA Res. 7 (2), 131-135 (2000)
20277480
2 (bases 1 to 92620)
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
Direct Submission
Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0842, Japan (E-mail:ynakam@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agg_graph.cgi?c=K14A17
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown: protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hobsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremli1.zozi.ilstate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MUI:5 and the 3' clone is MCE21.
Location/Qualifiers
1..92620
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="K14A17"
/clone_lib="Mitsui TAC"
complement(201..2180)

CDS
FEATURES
Source
```

```
/note="gene_id:K14A17.1"
/codon_start=1
/evidence="not_experimental"
/product="salt-inducible protein-like"
/protein_id="BAA94973.1"
/db_xref="GI:7670019"
/translation="MRGFASSRATATAAASKSLNASTSVNPKLSKTLNSSGKPTNP
LNORVISOVIERKDFLLINQBEFTTHRICLNTFVIVLQNDNPLHSEKRYLVKVS:
DPVYAKDOSLKSVLGNALFRGFLLSMELKAEIRDSGYRISDELMCVLISWRGLG
AYINDFAOISFLGKPKSTRYNADVALVKSNSLDLAYLKFQMRSDGCKPQFTY
NILHGVKKGVDEAIRLQKQMEQGNRPNVFTYITLIDFLGIAGRYDEALKLENM
RYVKLNQPLKATIGERYITPDSSTENAAKSCLLKGHDLVETCRIFDGVSRGVKPGFNS
MAKETGQNEATIGERYITPDSSTENAAKSCLLKGHDLVETCRIFDGVSRGVKPGFNS
YIVLVQALLNAORFSEGDRYLKQMGVDGLSSVYSYNAVIDCLOKARIENAAFLIE
MQDRGISPNLVTFNFTLSGYVRGDKVKVHGVLEKLLVHGKFPDVTLSLINCJCRA
KEIDAFDCFKEMLEWIEPNEIYINLIRSCSTGTDTRSVKLFARKKNGLSPLCY
AYNATIQSCFKRKYKAEELLKMLRIGLKPDNFTYSTLIKALSESRESEAREMEFS
STEREGCVPDSYTKRLVEELDKLSGLSREIVSAS"
join(3906..4124,4238..4462,4536..4649)
join(3906..4124,4238..4462,4536..4649)
gene_id:K14A17.2
similar to unknown protein"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAA94974.1"
/db_xref="GI:7670020"
/translation="MAKQDPTSSNTMLPLVSGSEKVRPOPLDJITGDTVKDAI:SNKVF
KVKPLFLGHNKRLVDPNDSPIVIMKKKVISKHQRWQYRGSDLDKFIIVKRSSTV
QKTRVEYFLKHNTRSCSDPTIKGRMKRACTIYVADSTKIIAQVYEGHERLVATI
YFNVDYATFIFLFDLIMMGTI"
complement(join(5031..5927,6011..6601,6984..7211))
/note="gene_id:K14A17.3"
/codon_start=1
/evidence="not_experimental"
/product="AMP-binding protein:"
/protein_id="BAA94975.1"
/db_xref="GI:7670021"
/translation="MAATKWRIDDLPKIPANYTALIPLWFLDRAAVY:PTPKSVIHG
SPREYTRQYDRCRLASALADRSIGPGSTVFVAILIAPNIPMYEAHGPVPMCGAVIN
CVNIRLNATVAFLSHSSOVSIMVDOEFTLAEDSLRLMEKAGSSKRPILLIVGD
HCAPESLNRAKSGAIEYEDFATGDPNIPQPPADEQWSIALGTSGITASPQGVV
LHRRGAYIMALSNPLIWQDQAVLYLTPMPCNGWCFPWSLAVLSGTSICIRQVTA
KEVSMIAKYVTHFCAPAVNLNAIYNAPEKDTILPLPHTVMTGAAPSPVFSFM
NOKGPRVATYGLSETSYSCVANKPEWDSLPPTQAKINARQGVRYTGMDQVIO
TOTGKVPADGKTAGENVIPRGNMVKGLKPNENKETFAGGHEHSGDIAVHPNYI
EKDRSKDYIISGENISVEVENYVYHHPAVLEASVYVAPRPERHQSCEKAFVILKSD
YEKDRQNKLAQDLMKFCRKLPAIVVPKSVFPGLPKTAIGKIQKHLTKAKEMGPV
PSRL"
complement(join(8463..8892,8977..9136,9230..9641))
/note="gene_id:K14A17.4"
/codon_start=1
/evidence="not_experimental"
/product="basic chitinase"
/protein_id="BAA94976.1"
/db_xref="GI:7670022"
/translation="MVSKPLFLSLLITVALVVFQGTGLVNAEDSPSSSTRKPLKVI
KGLKLGKGECKGWSEYCCNHTISDFETQYOFENLFSKRNSPVAHGVQMDYRSFIF
AAEYQPLGFGTAGELQGMKEVAFLHGVSKTSCGYGVATGGPLANGLCYKEMSP
DQYCDYKLYKLYPCTPGSYHGRGALPYWNYNGQTGEALXVLLSHPELENNAT
LAFQALRWMTTPPKKHLSAHDVFGKWKPKTNDTAARTPGFGATINVDYDQICN
SGFDNDMMNIVSHLYLIDLVGVEEAGPHEKLSAQDEPFSSSSAPSSGSSS"
join(13759..13822,14371..14445,14592..14707,14800..14888,
15014..15059,15238..15433,15520..15845,15928..17178,
17360..17713)
/note="gene_id:K14A17.5"
/codon_start=1
/evidence="not_experimental"
/product="transcription factor-like protein"
/protein_id="BAA94977.1"
/db_xref="GI:7670023"
/translation="MGDGLGRIGISEIHGFHTLQDLVQTMLEAKSKLNPNIHA
IUCGRILLDFDRKMLRNRKGNKKDKGTGRTKAHKLKVGNEIRTHVYAHGEDN
ITFVRRCYWLLKARENIVLVHYRDTQEAATTSQDSSISSPISVSEQTTFNVAADID
```

CDS

```

TVVRNEDLSLDINILDNWJELVPTCLNNNSAPTVDNLSYFTEPQLNNAANGTAERNSA
TVAGSILDALINDGQSGSEFGRWNSFISESGSLDEPSEFPMVMPQDDEIAPGVF
HSHNIPQVFNITVDSPAWAYSSKIKILVTFGLHDSYQHIFRNSLYCVAGGCFVFA
EYLQAGVYKRIIPSPGPNVNLISADGKHPIPLSGCFREHRAVPVLDKIVPEDNDPSK
WEEFEQVRLSHLLTSSNKNLVSLSKSPHNRDAKLSKNTLLHLLSWAYLVASIO
KNKVSFDAQKHLFSLSLKNRLKEMKVKVLPGRNTLDYDSKGLGVHLCASLGATWS
VOLFSLSGLSNFRKQGTALHWAAYVGRKNVAALISAGARPNVDSFKDNLGGC
MAADLAQNGYDGLAAYLAEKLCVQAQFDMKTAGNITGDLKACRKNLNGSLPDEEC
KRYDFFRKTEAARYIQCFQFTWKIRREYLNMRQAIRIOAFRGLCARROYKKILWSV
GVLEKAVLRKQKRGKRGFGLVAAEDSPGEAQEDFYKTSQRAEERLSRVVRQAM
PSKKAQODYRMKLTITHEAOVNHILFELNLSFGKKNRR"
complement(1015..18170..18172..18407..18490..18736..18807..
18977..19090..19168..19281..19387..19487..19590..19668..
19755..19805..19906..19989..20365..20431..20539..20669..
20916..21040..21126..21231..21345..21483..21624..21770..
21860..22172)
/note="gene_id:K14A17.6"
/pseudo
/codon_start=1
/evidence-not_experimental
/product="Lipoamide dehydrogenase"
join(29036..29206,29695..29826,29925..30009,30149..30312,
30830..31267)
/note="contains similarity to similar to ubiquitin
conjugating enzyme
gene_id:K14A17.7"
/codon_start=1
/evidence-not_experimental
/protein_id="BAA94978.1"
/db_xref="GI:7670024"
/translation="MADEYNKKNPAVKELIQVEKMQANSPDQFMSPLPLENIEHQ
FAIRPGDTEFGGIHGRIOUPADYPPRPMLITPNGRPEINFKICLSISNTHPE
HQPSVSTALVAGIAPFTSPNGALGSVDYKDPERRTLAIKRETPKYSUSPEROK
IIDETQYTSKATVYKPLPLECSQAPSVSSEHSQVEPQEAITVEERSLADTDC
VDQIIEFTAEAVNTAASVVAAPALPLPAVEVVVWVWVWVWVWVWVWVWVWVWVWV
WAAVLGTIAIMVLIILKFKKSNQSYSTGPMDDGS"
join(32260..32364..33121..33412,33506..33750,33837..34103)
/note="golAD26962.1
gene_id:K14A17.8
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAA94979.1"
/db_xref="GI:7670025"
/translation="MKSSNIVFDRIKERQMLSPFKIFQFADLSQCKRALPYEFTS
KYSEKDFSKMVIKQAGXITWVYSKNPTYYIETRGKDLFVSDNALQNEPFIETH
RGMVPHVNIYEONGLEMKPRKFTQMPGSSGIGKEEGENSLIDVKKESPESSPORA
EFLVRKKKTEDSKSKKMKTRNKVKKSKSKQKQVGVPEFKITTRKSYIKFLATIK
HFVCDHI PKSKSITFTRHPNGSKWKVCLVRLRITFISSGYSKLAREFELMGXDKTF
KLIKPEFVLLSKKRNKQDCMID"
complement(34528..34725,35213..35322,35660..35865))
/note="gb:AAF26101.1
gene_id:K14A17.9
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAA94980.1"
/db_xref="GI:7670026"
/translation="MAESGRRIGVAVDFSDSKKALSNAIONVVVDGRLILITIAH
DMNYEGEQLWETVCSPIPMSEFSDAAVMKYALKPDAETLDIVNTAARKKTIIVV
MKIYWDGPREKICAAAEQIPLSLVMNGRCLOGLKRMIMGVSYNHVVNVVACPVTVK
AHI"
complement(join(36309..36416,36556..36846,37250..37324,
37500..37634,38020..38145,38221..38374,38433..38573,
38878..38987,39134..39281,39474..39751,39993..40262))
/note="gene_id:K14A17.10
unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAA94981.1"
/db_xref="GI:7670027"
/translation="WDTNGCASLIEIGDQEEVEDPFLAFLDYARTVISPEDDEKEF
```

```

Query Match: 33.9% Score 576.2 DB 8: Length 92620;
Boot Local Similarity 99.5% Pred. No. 1.le-90;
Matches 578; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCGSGTATTATTGGATAACATGACAAAATATTTATTTATTTTCATGAGTTTATTTCATAG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71476 TGGGTTTTTATTTGGATAACATGACAAAATATTTATTTATTTTCATGAGTTTATTTCATAGTAG 71535

QY 61 CATGACAAATATTATATATATCATAGTGTATTAACATGTTTGTGTTCTTAAATAACATGCATT 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71536 CATGACAAATATTATATATCATAGTGTATTAACATGTTTGTGTTCTTAAATAACATGCATT 71595

QY 121 TTAATAATCAGACATTTGTTTTTAAATCAAAICTTAATCICTTTATATCAACACCATTCGAC 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71596 TTAATAATCAGACATTTGTTTTTAAATCAAAICTTAATCICTTTATATCAACACCATTCGAC 71655

QY 181 SGNAATTCAGGTAAAGAGAGAAATAAAGATGAGATAGAGATGAGATTCITATGGAATA 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71656 SGNAATTCAGGTAAAGAGAGAAATAAAGATGAGATAGAGATGAGATTCITATGGAATA 71715

QY 241 AGAAGAGAGACATGTAGTGAACAAAATAAAGAGATATGATGATATATTTTATGAGAG 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71716 AGAAGAGAGACATGTAGTGAACAAAATAAAGAGATATGATGATATATTTTATGAGAG 71775

QY 301 GTGTCACATTTATTTAGGAGAGGAGAGAGAAATAGAAAAGAAATGACATGTTGAA 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71776 GTGTCACATTTATTTAGGAGAGGAGAGAGAAATAGAAAAGAAATGACATGTTGAA 71835

QY 361 TCTGAAGAAGATGAATTTGTTTAAAGATCAAGAGAGAAAGAACTCCATGGCTAAAGTC 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71836 TCTGAAGAAGATGAATTTGTTTAAAGATCAAGAGAGAAAGAACTCCATGGCTAAAGTC 71895

QY 421 TGTAAAGAAGATGAAAAGAAACAAAAGAGAGAGAGAGAAAGAGAGCTTAAATAAGA 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71896 TGTAAAGAAGATGAAAAGAAACAAAAGAGAGAGAGAGAGAAAGAGAGCTTAAATAAGA 71955

QY 481 CTAACATATTGCCAAAATTTCTGTAGCCGACAAATCTATTTCGCCAAGGTTATTTCGTG 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71956 CTAACATATTGCCAAAATTTCTGTAGCCGACAAATCTATTTCGCCAAGGTTATTTCGTG 72015

QY 541 TATTCCTTTGAAGTCAAAAGTTATTTCTTACATATACTCTA 581
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72016 TATTCCTTTGAAGTCAAAAGTTATTTCTTACATATACTCAA 72056
```

```

RESULT 3
AB005248/8
LOCUS 83646 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, PI clone MX110.
ACCESSION AB005248 BA000015
VERSION AB005248.1 GI:2264320
KEYWORDS
SOURCE
Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui PI
clone:MX110.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (sites)
AUTHORS Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M.,
Miyajima,N. and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. I.
Sequence features of the 1.6 Mb regions covered by twenty
physically assigned pi clones
JOURNAL DNA Res. 4 (3), 215-230 (1997)
MEDLINE 97471969
REFERENCE 2 (bases 1 to 83646)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1997) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp.
```

COMMENT

tel:81-438-52-3935, Fax:81-438-52-3934)
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/aga_graph.cgi?c=MX110
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Graft
 (Informatics Group, Oak Ridge National Laboratory,
<http://combio.ornl.gov/Graft-1.3/>),
 GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>),
 NetGen2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, <http://www.cbs.dtu.dk/services/NetGen2/>) and
 SplicePredictor (Volker Brendel, Stanford University,
<http://gremli.ni.zool.iastate.edu/cgi-bin/sp.cgi>).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is MX117 and the 3' clone is MX8-B.

FEATURES

source

1. .83646
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MX110"
 /clone_lib="Witsui Pl"
 join(1791..2912,3016..3318,3584..4560)
 /note="gene_id:MX110.1"
 /codon_start=1
 /evidence=not_experimental
 /product="disease resistance protein-like"
 /protein_id="BAB09346.1"
 /db_xref="GI:9758812"

CDS

/translations="MTEKATDVSNMLNYSRSDFDGLIGQARMKEMESLLCLDS3
 EVRMIGSPGIGKT...ARVLYSOFSENFFELSIEMGNIKELMTKTRPVCSEYSAKQ
 LQKQFSLQINKHMDLHLHGVAGQRLNCKKVLVLDSIDQSIQDIAKATKWFQHG
 SRLIITQOKLLKANGINHIYKVEPPSYAYQMFQKARQGNPNQDFEELAKVFI
 KLGLHPLGRVNGSHFQRMGSFVWVNALPRUKIRIDASTISILKSTFALCEKRL
 FLIIACLFNQNKVEVDYLAUSELDVROGFHLAEKSLINKLSTICTEIMENLL
 VQSGIDVFRKQHQICCPGRQFLDIDARDICEV.TDNIQNRNVVGIFFLEVRNLSC
 LNTSEAFQCMNMLKPRPHQDDESDKLYLPQGLNLPQKRLIENSREPMTCLFS
 NCTGYLYVEIRKMSKLNQWGNQPLGKRMDSKSLKELPDLSTANLNYELIM
 ISNIDKLITATKVEVSTIKSHLRKLEKSYSENKSEPHALD:YITLYINDPE
 MQPIQWYKXISHLOTLEGCKRLVTIPQLSQSLQVWVINGCESLERLNSFQNHPE
 RFUWFNCFKLNEAREFTQSTHAILPSREVPAFTFRANGSSIMVNLNHRPLSTI
 LRFKACVLLYKRIIDNDKEAADRTVTPRIRENDKIGVDVQVWYRFHVPTILEDHLL
 TFQIEADVTSNKLIFSFDIKNNQAVIKBCGVLO."
 complement(join(5504..5620,5688..5736,5836..5905,
 5990..6122,6206..6250,6374..6688))
 /note="gene_id:MX110.2
 pir|IT03994
 similar to unknown protein"

CDS

/codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09347.1"
 /db_xref="GI:9758813"
 /translation="MMASSLSPFCPLSYEKVPIAAATNSPATTIPYKXKXGOTRIHL
 KIPSGLKMEVIEQPKSKRENPQVVFV:SSVFAAKWAENKLPFFSSSGTDSYAVSL
 LGGSEDEPLCTVAGTLOTHASDIADPTESNLGSSPPVLVGHSGGL:VOYLIANIN
 KSLGTENAPELMSGAVMCSVPFSGNSGLVLR:LFSKPVAAFKVILSLAAKGFQKST
 PLCRETFFSQMDQLVKRYNLR*
 join(7788..7835,8043..8777,9032..9157,9205..9901,
 10085..12958)
 /note="gene_id:MX110.3"
 /pseudo
 /codon_start=1
 /evidence=not_experimental
 /product="non-LTR retroelement reverse transcriptase-like
 protein:"

CDS

complement(13525..13827)
 /note="gene_id:MX110.4
 pir|IT02618
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09348.1"
 /db_xref="GI:9758814"
 /translation="MKQVILVIYSLIKMTKLRSCLDCNSSTYITQLKHLKNEHUG
 LVLLWVQSQMRCCEGMKRVHYVDLMSATEPHIAEVVHQLPYLDTLILFPCIS*"
 join(18792..18841,19226..19412,19494..19579,19875..19953,
 20464..20507,20697..20783,20894..21006,21097..21162,
 21216..21308,21350..21509)
 /note="gene_id:MX110.5
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09349.1"
 /db_xref="GI:9758815"
 /translation="MNSPKQISDQSGSHRHHPFQIIHLIGNFFRIKWSYKIRVLN
 QTGAPVULFECCLVPSSIFLIIQKPKGRALSNQOIVPSLINGVYITLITLW3KG
 LKSCGPLRAILSEYSGAVLGVSLVGRGHVMMKLPTKSDTETKEEVEYQTEAL3KG
 M3MIPFAGILSALRRVIARVSKKQOKRLHAIITISATCFLEFPVAMKDIIGSS
 SCKTSELPEFAWFLSTIIIGIILIFVYDNIABERLHMVFPSPRLMWACVCIIVDI
 RVNIRK3ISISTEORLEFVEIRSIKRNPCSPFIPYINDVLLNLVLDVNL3GFIK
 FSIYVPSFFCFGLGERAYANLV*"
 join(26073..30268,30416..31552)
 /note="gene_id:MX110.6"
 /pseudo
 /codon_start=1
 /evidence=not_experimental
 /product="retroelement pol polyprotein-like"
 32292..33620
 /note="contains similarity to retroelement pol
 polyprotein
 gene_id:MX110.7"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09350.1"
 /db_xref="GI:9758816"
 /translation="MVETKRIKRSKTMSTEPQSEPAVHTKEDKKALKVISP
 SEPTFAREILEKMGJLETDIEFTKNGLAFTSMDDHTFSEPTVDFSTMEYTTIN
 PPIAKSGI:NFKVTKAFSISIPDCEAYGFENKSEMFPPKFGIHFVWELASGK
 F3CNASIKSRVPIRYALKLHALFORGETSTTVSEMCFFQGVKELLEDADEN
 GGRVLHDDGNDVNYGICFASYLEDKNRILKIGKTSIV*GGTLPFLVKAGVDLSP
 FNALPKRYIDYENLVRSCLKRSDDAFIYLYTRDNKLECLILPSKEIKVESHEDI
 MFLPSKQVQNLQGMLEIEDLPPSPROPYLNKCDLPSTFEPDKTKMELQSV
 IKSQRKIKKSYAIGKLLKKVKLQPPDYVSSDELDLVFFSNEEDRDLTESDENNE
 TPAISHSRNRKASRHSYMP*"
 join(36310..37248,37409..37549,37633..37824,38363..38519)
 /note="embicAB86928.1
 gene_id:MX110.8
 similar to unknown protein"
 /pseudo
 /codon_start=1
 /evidence=not_experimental
 /join(39564..40460,40627..40767,40853..41193)
 /note="db:AAD55600.1
 gene_id:MX110.9
 strong similarity to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09351.1"
 /db_xref="GI:9758817"
 /translation="MDLLSNLPYELLCHLSFLT:KEAALTSVLSKRWRLIAFVPRV
 DIVDHDIRELDFDVRVALQGSNPINKFSLDCSGVSDSDVDCVQINWVRGVSEIN
 L5IFVDSVFEDDYNLFPKVFENKVLKGLSYISWLOGSIFLPMKLVLESVLSVE
 KFEILLHALPALLEELVMNNIYKSRDQVVDLIVSSESLKTLTINFIYVTHLTSFDIPI
 AYLSVSGYVDYDAEAKMENLPEARISLLYVEGDISVRALINNDLEDEYDLOPE
 NVKLMJNIRICLYLSPNTLEVLCCESMPVPVKNLKSLSTKASNRWQAMPVLL
 RNCPHLETVLLEGLLHVYDKGACDCVSRDKRSLTSCPVKYLEIKGFGTTKEM
 NMIKHFLEVPCLKELKMYMEENDPAQURVPEVITGEMMYRYKSSSCNVQLLVGKGL
 YKKWtp"

```

CDS      join(42817..43766,43859..43975,44083..44493)
        /note="emb|CAR86928.1"
        gene_id:WX110.10
        similar to unknown protein"
        /pseudo
        /codon_start=1
        /evidence=not_experimental
        4015(45089..45677,45843..47158,47312..47452,47544..47871)
        /note="contains similarity to non-UTR retroelement reverse
        transcriptase
        gene_id:WX110.11"
        /pseudo
        /codon_start=1
        /evidence=not_experimental
        48978..49433
        /note="gene_id:WX110.12
        unknown protein"
        /codon_start=1
        /evidence=not_experimental
        /protein_id="BAB09352.1"
        /db_xref="GI:9758818"
        /translation="MHVDSMKRARSVAVEGKERSSNOPTREOSKHAKTAKAR
        MNVAKSMPLPKPKKSLMILSMKXEDTODDQNNVTVVEKEKSEKQTVQNTRE
        NAEMLSKVSEVPHWLYEEBWTWFAFGHGWQRYDAVMGSKSFN"
        complement(join(50633..50972,51043..51177,51269..51438))
        /note="gene_id:WX110.13
        sp|P10798"
        /codon_start=1
        /evidence=not_experimental

Query Match      31.8%; Score 540.2; DB 8; Length 83646;
Best Local Similarity 93.6%; Pred. No. 26-84;
Matches 574; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

y 1 TGGGTTTTATGATTAACAGACAATAATTTATTATTCAGAGCTTTTATTGAGAG 60
b 23771 TGGGTTTTATGATTAACAGACAATAATTTATTATTCAGAGCTTTTATTGAGAG 23712
y 61 CATCAGAAATTAATATATCATGCTGTATATACATGCTTTTGTTCTTAATATCATGCGAT 120
b 23711 CATCAGAAATTAATATATCATGCTGTATATACATGCTTTTGTTCTTAATATCATGCGAT 23652
y 121 TTAATAATCAGACATTTGTTTAAATCAAAATCTAATCTCTTATATACAGACATTTGAC 180
b 23651 TTAATAATCAGACATTTGTTTAAATCAAAATCTAATCTCTTATATACAGACATTTGAC 23592
y 181 GSAAAATTCAGTAAAGAGAAATTAAGAATGAGAGATGAGAGATTTCTATGSAATA 240
b 23591 GSAAAATTCAGTAAAGAGAAATTAAGAATGAGAGATGAGAGATTTCTATGSAATA 23533
y 241 AGAAGAGAGAGAACATGTAGGTGACAAATAAAGAGATATGATGATATTTATCAAG 300
b 23532 AGAAGAGAGAGAACATGTAGGTGACAAATAAAGAGATATGATGATATTTATCAAG 23473
y 301 GTGGTGAAGATTTATTAGAGAGGGAGAGAGAAATAGAAAAGAGAAATGACATGCTGAA 360
b 23472 GTGGTGAAGATTTATTAGAGAGGGAGAGAGAAATAGAAAAGAGAAATGACATGCTGAA 23415
y 361 TCTGAAGAAGATGAATTTGTTTAAAGATGAAGAGAGAGAGAACTCCATGGCTAAAGTC 420
b 23412 TCTGAAGAAGATGAATTTGTTTAAAGATGAAGAGAGAGAGAACTCCATGGCTAAAGTC 23354
y 421 TCGTAAGACATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
b 23352 TCGTAAGACATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 23293
y 481 TTAATTTGCAAAATTTCTGTAGCCGACAAATTAATTTTGTCTCCAAAGTTATTTCG 540
b 23292 TTAATTTGCAAAATTTCTGTAGCCGACAAATTAATTTTGTCTCCAAAGTTATTTCG 23233
y 541 TATCTTTTGAAGTCAAAAGTTATTTCTTACATATATCTTAAATAATAGCGATACCAA 600
b 23232 TATCTTTTGAAGTCAAAAGTTATTTCTTACATATATCTTAAATAATAGCGATACCAA 23173

```

```

Q# 601 TTTTCCACACAT 613
| | | | |
DL 23172 TATTTTAAATAT 23160

```

```

RESULT 4
AC012394
LOCUS
DEFINITION

```

AC012394 94487 bp DNA linear PLN 12-OCT-2000
Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence,
complete sequence.

```

AC012394
VERSION
KEYWORDS
SOURCE
ORGANISM

```

Arabidopsis thaliana.
Arabidopsis thaliana.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

```

REFERENCE
AUTHORS

```

Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
Wu, D., Mall, R., Rong, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence
Unpublished
2 (bases 1 to 94487)
Lin, X. and Kaul, S.
Direct Submission
Submitted (27-OCT-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

Lin, X. and Kaul, S.
Direct Submission
Submitted (10-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
4 (bases 1 to 94487)
Town, C.D. and Kaul, S.
Direct Submission
Submitted (12-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

Lin, X. and Kaul, S.
Direct Submission
Submitted (05-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

Lin, X. and Kaul, S.
Direct Submission
Submitted (05-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

Lin, X. and Kaul, S.
Direct Submission
Submitted (12-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

Lin, X. and Kaul, S.
Direct Submission
Submitted (12-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

Lin, X. and Kaul, S.
Direct Submission
Submitted (12-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

Lin, X. and Kaul, S.
Direct Submission
Submitted (12-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

Lin, X. and Kaul, S.
Direct Submission
Submitted (12-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

Lin, X. and Kaul, S.
Direct Submission
Submitted (12-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

Lin, X. and Kaul, S.
Direct Submission
Submitted (12-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
Simple repeats are identified by repeatmasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

```

source
1..94487
  /organism="Arabidopsis thaliana"
  /cultivar="Columbia"
  /db_xref="taxon:3702"
  /chromosome="1"
  /clone="F15M4"
  complement(967..1017)
  /rpt_family="AT-rich"
  complement(1289..1338)
  /rpt_family="AT-rich"
  complement(14053..14080)
  /rpt_family="AT-rich"
  complement(6943..9510)
  /gene="F15M4.1"
  /note="identical to hydroxy methylglutaryl CoA reductase
  (AA 1-592) GB:X15032 [Arabidopsis thaliana]"
  complement(join(<6943..7515,7507..7953,8115..8296,
  8522..>9510))
  /gene="F15M4.1"
  complement(join(7255..7515,7507..7953,8115..8296,
  8522..9510))
  /gene="F15M4.1"
  /codon_start=1
  /product="hydroxy methylglutaryl CoA reductase (AA 1-592):
  9510-7255"
  /protein_id="AAF16652.1"
  /db_xref="GI:6554470"
  /translation="MDLRPRPKPPVNTNNSNGSFSSYOPRTSDCRRERNTIAP
  FKASALPLPLYNLNAVFFLFFSVAYLLHWRDKIKYNTPLHVYIIIEGALIALI
  EPVIFLLGFGFQVQSIRASQANDLAUTIDDDHRLVTCSPPTFLVSVAKLPP
  BPIVTSPEDEETVKSIVDGVIPSVLSLEKGCXRAASIRKALQRTVGRSIEGL
  PLDGDYFSLGQCCMPVGIQIPVGIAGPLLDGVEYSVPMTTECCVLASTNRCG
  KAMFLSGGATSTVLDKGMTAPVVRFAARRASBLKFLFENPFDILAVVFNESSPE
  ARLOSQKTIAGKNAYVRFCSTGDAMGMNVSKGVNLEYLTDPDFMDVIGISGN
  PCSDKKPAVNVKIEGRGSVCEAVIGEIVNVKLTSVRALVELNKLNLASAVAS
  SLGGNNANSTVSVFATGQDPQNVSSOCIMNEANGDKDHIISVTMSIEVCG
  TVGGTGQASQACNLGIVGASTSPGMNRRRLATIVAGVLAGELSLMSATAACQ
  LVRSIMKYNRSRQISGATTTTTTT"
  complement(7258..7283)
  /rpt_family="(CAA)n"
  complement(11154..11209)
  /rpt_family="AT-rich"
  complement(11551..11580)
  /rpt_family="AT-rich"
  complement(11933..12004)
  /rpt_family="AT-rich"
  12390..13268
  /gene="F15M4.2"
  /note="similar to unknown protein GB:AAD30653 [Arabidopsis
  thaliana]"
  join(<12390..12391,12491..12629,12730..12861,12929..13019,
  13099..>13268)
  /gene="F15M4.2"
  join(12390..12391,12491..12629,12730..12861,12929..13019,
  13099..13268)
  /gene="F15M4.2"
  /codon_start=1
  /product="unknown protein: 12390-13268"
  /protein_id="AAF16653.1"
  /db_xref="GI:6554471"
  /translation="MFDVKALVVGKELWRETIPLCSKSKYKLGSLKSNSEYEVKISY
  PASIPARVSICLKNREVELKNOMRLNTEKCFKAFESKEVNNKAGLNVLVLEP
  EGVALPNSRRFVILVIVCEQLMGIPYSSWSVLLVALCLVSVFVPSRLPSSLI
  TRQQLNSHRSHGKDS"
  14056..15506
  /gene="F15M4.3"
  /note="similar to cinnamoyl-CoA reductase GB:CAA56103
  [Eucalyptus gunnii]"
  join(<14056..14170,14277..14431,14516..14701,14819..14990,

```

```

15096..15285,15362..>15506)
  /gene="F15M4.3"
  join(14056..14170,14277..14431,14516..14701,14819..14990,
  15096..15285,15362..15506)
  /gene="F15M4.3"
  /codon_start=1
  /product="putative cinnamoyl-CoA reductase; 14056-15506"
  /protein_id="AAF16654.1"
  /db_xref="GI:6554472"
  /translation="MAYQKVCVTGAGGFIASMLVKFLLSRGVTVHGIVRUCDEKND
  HLKLDNASKNLKLFKADLPDDEGLNSIDGSGVFHIASPVPEGELIKPALGTGK
  NVLEACTETKVKVYVYSSIAAVYNFNPQDVAKEDCSQIYDLSEGIWRYIYL
  AKTLTERALEMSKRNFAVYVTLCPSTVIGPRLSTLNSSLGLKFKIKGGIKSLISD
  ELYLVYDVADALLVYENREATGICNSHSYIDSLMEKKNMYPKRNFPESFTE
  VKKEVRPLSAEKLKLNKGKFLPELTIDTSVSWFEAAGDLPKA"
  16281..16361
  /rpt_family="(CAT)n"
  15955..18423
  /gene="F15M4.4"
  /note="Contains Pfam profile: PF00076 RNA recognition
  motif. (a.k.a. RRM, RBD, or RNP domain)"
  join(<16955..17161,17385..17529,17639..17755,17844..17884,
  17978..18215,18332..>18423)
  /gene="F15M4.4"
  join(16955..17161,17385..17529,17639..17755,17844..17884,
  17978..18215,18332..18423)
  /gene="F15M4.4"
  /codon_start=1
  /product="putative RNA-binding protein: 16955-18423"
  /protein_id="AAF16655.1"
  /db_xref="GI:6554473"
  /translation="MAYQVPVSGFHYLNSPFGDTTKVTVGGLAMPTQSETLRQHF
  EYGEILCAVVIADKNTGSKGYGFVFRDEPAARRACADPTPIIDRRANCLASLG
  RRPPLPFAVNPMDLLPRILEMCRVLEHLEIRISNHFIIITSKELSLCHVTA?
  GREYVNSQGLYSPYMQQYLQVYGVGAVNSPVYQGLSOTIPNGHGTAVQVYS
  GSHILQLGPTVSTMTTSSMPALQAPYPSGIRGPAPVQSHIIVHSPQMGSTASQC
  "TR"
  complement(19352..20843)
  /gene="F15M4.5"
  complement(join(<19352..19378,19455..19493,19581..19699,
  19777..19852,19942..20024,20291..20356,20433..20536,
  20615..20690,20777..>20843))
  /gene="F15M4.5"
  complement(join(19352..19378,19455..19499,19581..19699,
  19777..19852,19942..20024,20291..20356,20433..20536,
  20615..20690,20777..20843))
  /gene="F15M4.5"
  /codon_start=1
  /product="unknown protein: 20843-19352"
  /protein_id="AAF16656.1"
  /db_xref="GI:6554474"
  /translation="MAAISPWLSSPQSFNPRVITIDSRCCSSISAAISVLDSNEEO
  HRISRDHVGMKRRQVLMQIASSVFFLPALSPAFETNASEAFRVYTDSTNKFESI
  PVSVAITGLGPDFTRMESFGKVPAPATLYSLGDRSQKVPQVGTAKLIDSRASKGY
  FYEITLQNPGEARKHLYSAIGMATNGWYNRLTYVTGTFDDESAEQSKIKQTVKSPR
  FI"
  complement(20998..21026)
  /rpt_family="AT-rich"
  complement(21140..21167)
  /rpt_family="AT-rich"
  22462..23090
  /gene="F15M4.6"

```

```

Query Match      31.4%; Score 534.6; DB 8; Length 94487;
Best Local Similarity 96.5%; Pred. No. 1.8e-83;
Matches 557; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

Cq 1 TGGGTTTATTGCAATACATGACAAATATTTATTATTTCATGAGTTTTTTTTCGATAG 60
      |||||
Cb 1317 TGGGTTTATTGCAATACATGACAAATATTTATTATTTCATGAGTTTTTTTTCGATAG 1976
      |||||
Cy 61 CATGACAAATATTAAATATCATCGTGTAAATACATGTTTGTCTTAATATCATCGATT 120
      |||||

```

2b 1977 CATGACAAATATTAATATATTAATGTTAAACATGTTTCTTCTTAAATATACATGCAAT 2036
 4Y 121 TTAATATCAGACATTTGTTTAAATATCAATATCAATCTATATATACAAACAGACATGAC 180
 2b 2037 TTAATATCAGATATTTGTTTAAATATCAATATTAATCTTCTTATATACAAACAGATGAC 2096
 2Y 181 GGAATATCAGGTAAAGAGAGAAATTAAGATAGAGATAGACATAGACATGATGATGATGATG 240
 2b 2097 GGAGATTCAGGTAAAGAGAGAAATTAAGATAGAGATAGACATAGACATGATGATGATGATG 2156
 2Y 241 AGAAGAGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 299
 2b 2157 AGAAGAGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2216
 2Y 300 GTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 359
 2b 2217 GTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2276
 2Y 360 ATCTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 419
 2b 2277 ATCTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2336
 2Y 420 CIGCTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 479
 2b 2337 CIGCTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2396
 2Y 480 ACTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
 2b 2397 ACTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2456
 2Y 540 GTATCTTTTGAAGTCAAAAGTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 576
 2b 2457 GTATCTTTTCAACATCAAGGTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2493
 RESULT 5
 C015450/C
 OCS
 EFinition Arabidopsis thaliana chromosome 1 BAC F1465 genomic sequence.
 CESSION AC015450.5 G1:12323968
 ERSION
 EYWORDS
 OUNCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops-s.
 Lin.X., Kaul.S., Town.C.D., Benito.M.-I., Greasy.T.H., Haas.B.J.,
 Wu.D., Maiti.R., Roeding.C.M., Koo.H., Fujii.C.Y., Utterback.T.R.,
 Barstead.M.E., Bowman.C.L., White.O., Nierman.W.C. and Fraser.C.M.
 Arabidopsis thaliana chromosome 1 BAC F1465 genomic sequence
 Unpublished
 2. (bases 1 to 100806)
 Lin.X. and Kaul.S.
 Direct Submission
 Submitted (16-NOV-1999) The Institute for Genomic Research, 9712
 Medical Center Dr. Rockville, MD 20850, USA, xlin@tigr.org
 3. (bases 1 to 100806)
 Town.C.D. and Kaul.S.
 Direct Submission
 Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr. Rockville, MD 20850, USA, cdtown@tigr.org
 On Jan 19, 2001 this sequence version replaced g1:12280761.
 Address all correspondence to:at@tigr.org

BAC clone F1466 is from Arabidopsis thaliana chromosome 1
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.
 Genes were identified by a combination of several methods: Gene
 prediction programs including Genscan+ (Chris Burge,
 http://CCR-081.mt.edu/GENSCAN.html), GenemarkEM (Mark Borodovsky,

http://genemark.biology.gatech.edu/Genemark/), GlimmerA (a variant
 of GlimmerM, see Mihaela Pertea,
 http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and
 Geneslicer (Mihaela Pertea and Steven Salzberg, contact
 mperle@tigr.org), searches of the complete sequence against a
 peptide database and the plant EST database at TIGR
 (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
 indicate the level of evidence for their annotation. Genes with
 similarity to other proteins are named after the database hits.
 Genes without significant peptide similarity but with EST
 similarity are named as unknown proteins. Genes without protein
 or EST similarity, that are predicted by more than two gene
 prediction programs over most of their length are annotated as
 hypothetical proteins. Genes encoding tRNAs are predicted by
 tRNAscan-SF (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SF/).
 Simple repeats are identified by RepeatMasker (Arian Smit,
 http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES

Location/Qualifiers
 1..100806
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="F1466"
 misc_feature
 1
 /note="35000 nt before this point were not included in
 the submitted sequence, due to overlap with another BAC
 [F15M4]."
 5217..5774
 /gene="F1466.1"
 /note="contains Pfam profile: PF00097 zinc finger, C3HC4
 type (RING finger)"
 5217..5774
 /gene="F1466.1"
 5217..5774
 /gene="F1466.1"
 /codon_start=1
 /product="putative RING zinc finger protein; 5217-5774"
 /protein_id="AAG51946.1"
 /db_xref="GI:12323975"
 /translation="MARLLFLLQEA NSTSPAESPPFNSDLVLAVLCAITCIIG
 IAVSRCAWLRIASNRSDOTHPVPVAAANKGLKVKVLSLPKLTYSPOSPFAKLV
 PCAICCTEFAAGDELRLVLPQCCHGFGVSCIDTILGSHSSCPGCRJLIVIRCHKQGL
 PGSSSGPEPPTRIKQREDGPUNLP."
 complement(8331..10137)
 /gene="F1466.2"
 /note="N-term similar to N-term of NAM GB:CAA63101
 [Petunia x hybrida] (apical meristem formation), CUC2
 GB:BAA19529 [Arabidopsis thaliana], GRAB2 protein
 GB:CAA09372 [Triticum sp.]"
 complement(join(<8331..8855,9526..9800,9933..>10137))
 /gene="F1466.3"
 complement(join(8331..8855,9526..9800,9933..10137))
 /gene="F1466.2"
 /codon_start=1
 /product="unknown protein; 10137-8331"
 /protein_id="AAG51953.1"
 /db_xref="GI:12323982"
 /translation="MMLAVEDVLSLAGEPNERNEGLPPGFPHPTDEELTFYLASKI
 FHGGLSGIHISEVDLNKCEPHELMEMAKGERWYFYSLDKRPYPTGLRNRRATYAG
 WKATQKREVFSGGGQGLVGMKTLVYKGRAPGLKWKVMHEYLENDHSHRHTCK
 FEWTCRVENKTGRKNVGLIHQISYLHNHSLSTTHRRHEALPLLIEPSNKLTFNF
 PSLYDDPHQNNNNFLHSSGHNIDELKALINPVVSLQGLIFFSGNNNDGDFD
 FNLGVKTBQSSNGNEIDVRYLENPLFQEA SYGLGSSSSPGPLRLMLDSCPGLGQL
 ."
 complement(15417..18176)
 /gene="F1466.3"
 /note="similar to phosphate transporter GB:BAA20522
 [Catharanthus roseus]; contains Pfam profile: PF00083
 Sugar (and other) transporter"
 complement(join(<15417..16305,17467..>18176))
 /gene="F1466.3"
 complement(join(15417..16305,17467..18176))

ORGANISM	Arabidopsis thaliana	thaliana Chromosome II and IV
REFERENCE	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: curculas ii: Brassicaceae: Brassicaceae: Arabidopsis.	contains EST gb:T04634, T20404, T20458, T46059, Z17562"
AUTHORS	1 (bases 1 to 96453) McCombie R.W., Spiegel L.A., Huang E.N., Nasolmento, L.U., de la Bastide, M., Vil, D.M., Preston, R.R., Marero, A., Shah, R., O'Shaughnessy, A., Rodriguez, M., Shekter, M., Schutz, K., See, L.H., Saby, J., Habermann, K., Dedhia, N.N., Kewes, H.W., Lemcke, K. and Mayer, K.F.X.	/codon_start=1 /product="putative protein" /protein_id="CAB82107.1" /db_xref="GI:7321060"
JOURNAL	Unpublished	/translation="MPLKKRKGVTGTVKSTTEETATATKETAPATKATPATKEITA
REFERENCE	2 (bases 95732 to 104386) McCombie R.W., Robson, J., Grycmanprez, B., Bastiaens, L., Volckaert, G., Kewes, H.W., Lemcke, K. and Mayer, K.F.X.	PIITETAPTKETATPATETATRTREESLTDQSPENVEESEESEEESEEESEEESEEE EAGEEEEEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEE CMNECLIQEIDEETEAAPPLSMYPPSEYVKKIKLSTRCYTHELLTTFDKLEPMSK SERLWFENHPSQHFHMPDPNHRMLGMWMLLRTARIEKKEAWFTVNGVPIRYGI LEHALISPCNCKYKNTGNLDPKRRKFDVVKREDVREKIGMVPEGESKER LRMVLYELSSJIIAPIKTKDAPQVDFELKVMSDLJFCRNFONRGYSEDMVLTGS HTVNHFGSVTNNKXIYIPVPGFCFLMELLAFEAISQLREKFIETIADADGCCPMCK VFKKXHLKGFPLDTIYAEGLTIQVDSIVTLGIEQKLMKAITEEEDVDVADVMVL SWEKKRKLRLFDQLQODIQALQIAGPIGAEDIPAVRGQELPSLVSVVOMIE NLDKRWDMNNKIIIMISDLRRVESLAFKDEQKAKEEKVHIDNCVILSNMTIRNQ DMNKEEAGDSREKQDEEDVQEKDQDEEDVQEKDQDEEDVQEKDQDEEDVQEKDQDE EKEKEDPKKDGSPKPEANVTTRITRTKN"
AUTHORS	Direct Submission	24970..25451 /gene="At4g08710"
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lencke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de	/number=1 25452..25914 /gene="At4g08710"
COMMENT	On Mar 23, 2001 this sequence version replaced gi:5103797. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thai/ .	/number=1 25915..26692 /gene="At4g08710"
FEATURES	location/Qualifiers	/number=2 26693..26785 /gene="At4g08710"
source	1..104386 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4"	/number=2 26786..26947 /gene="At4g08710"
misc_feature	1..22908 /note="Overlap to BAC T3F12; please refer to EMBL:AC002983 or analysis and annotation"	/number=3 26948..27033 /gene="At4g08710"
gene	22865..24293 /gene="At4g08700"	/number=3 27034..27477 /gene="At4g08710"
gene	complement(join(22865..23096,23227..24293)) /gene="At4g08700"	/number=4 27478..27516 /gene="At4g08710"
CDS	complement(join(22865..23096,23227..24293)) /gene="At4g08700"	/number=5 27519..27800 /gene="At4g08710"
	/note="Similarity to various hypothetical proteins of Arabidopsis thaliana Contains Prokaryotic membrane lipoprotein lipid attachment site Aa82-92; Prokaryotic membrane lipoprotein lipid attachment site Aa235-235 contains EST gb:AI99722.1" /codon_start=1 /product="putative protein" /protein_id="CAB82106.1" /db_xref="GI:7321059"	28303..31953 /gene="At4g08720" complement(join(28303..28494,28565..28761,29626..30162,30333..31953)) /gene="At4g08720"
exon	complement(22865..23096) /gene="At4g08700"	complement(join(28303..28494,28565..28761,29626..30162,30333..31953)) /gene="At4g08720"
intron	complement(23097..23226) /gene="At4g08700"	/note="similarity to hypothetical protein F2686.15 - Arabidopsis thaliana, PID:g3242713"
exon	/number=1 complement(23227..24293) /gene="At4g08700"	/codon_start=1 /product="putative protein" /protein_id="CAB82108.1" /db_xref="GI:7321061"
gene	24970..27800 /gene="At4g08710"	/translation="MVYKVIYFRMKDRAYSLMNGKTIITLSMLESRICSKGLDESKY KLQMKYNVAMGLGTQIHIISDDEGVYISSGMSVSRVPLTVDTVPPELPEE:SRVE VNMKSSIGKNGYDGLSGNDEMRDNATIVIGELDGOGTELGAIIVVDDTQDHEISGRD DRDDDRGDEYVEPPDVESSQFKKEWEDIGLTLRQEPFSAALQEVYDQKAFANSF DIVIKKSKERTVLTCSCANDWRIOACVQDTISITRRYNNKSHSRISKSITMR KRKGTPELVAAHLHTDTPGLLETPPKVIMELVQIKLVKYSYHSLRQKQAVCDLK GSAEDSYKIDNGYLYMKRVKNDGTVMYKLDGSKGFQVLYFIALGASIEGFQAMKVI VDATHLKNYGGVLYFASARDPNRHHYIIAVGLVDGENDASMGWFFELKLSVPTDPE LYFMSDRNSLLKGINATYAAHGYCWHLSQNVKHAHSTNINRDYLAHWFNELSRYY TWSEPREYRKYRPTAVNTLEDSTVKEKWSGVNDKFPVPLEVNLHNLWDAEKLK VTENLNTFRLEYNVRGVEGNIYLVNLFKTSCKQVFDIQKPCIHAAAFIVFSDTTR TGMELHELVSRYWAEALWALAYGTIYLVGRSQWDIPDDIKALKVLPKPKKKGR TKVLRIPSTGERPKRQRTQNKRRPQSLQWLLFGTTRMDRLCESDPWYGEKMSAKRI
CDS	complement(24970..25451,25915..26692,26786..26947,27034..27477,27519..27800) /gene="At4g08710"	
	/note="similarity to hypothetical proteins of Arabidopsis	


```

/gene="AT4g08620"
/notes="similarity to various hypothetical proteins of
Arabidopsis thaliana Chromosome II, III and IV"
/codon_start=1
/product="putative protein"
/protein_id="CA878007.1"
/db_xref="GI:7267525"
/translation="MSVSNKSGVGGGILKFNVIDAFVRERVRLFPNCEKGPVIT
IGEEVLEAMNIGAKRMKVIKFFPIRLPELEEYLAALTGSGMRVFGKYLIVCGNSPFD
PLDDAVTTPVYRLNIPVNYHRLLEBEIAGLKLKVDLNIITFGGRFARVCZ
EVNLAKPLGTVLINDRYEYVAEEVEEGTIVVQSGIRVEKPAQKAFVAAAGSGRS
NRSLEPLKNGQVEIENRFGLEDVMSAFIGEVAILEGSNKENKYGKXMKRESGVIT
QVKEACQVITFKKGCKGKNGKPKKRWKGLIIMEPAGPRCKQGSVKNKPVKGLTFGPIK
GESSNLVGDAISANGKRLRVEQSVGRPGGAFVSGLETRSGEMKXSMELPQNNHQQC
SKELLCDTAIVV"
exon: 5932..9093
/gene="AT4g08620"
/number=1
exon: 9094..9192
/gene="AT4g08620"
/number=1
exon: 9193..9553
/gene="AT4g08620"
/number=2
exon: 9554..9685
/gene="AT4g08620"
/number=2
exon: 9686..10347
/gene="AT4g08620"
/number=3
gene: 10891..11438
/gene="AT4g08630"
CDS: join(10891..11025,11244..11393,11430..12120,12151..13089,
13162..13585,13634..14348)
/gene="AT4g08630"
/notes="similarity to various putative reverse
transcriptases of Arabidopsis thaliana
Contains Anticacy-1 transfer RNA synthetases class-II
signatures AA699-708"
/codon_start=1
/product="putative protein"
/protein_id="CA878008.1"
/db_xref="GI:7267526"
/translation="MSPDSLEFGIEWINRVSLDMGTSGNKYTKRGRVENTVAKRLD
RLAKNRDVGDKKKEEMREIVVOADLVQSDMLKKEEDLIKAFDKSREKWI
ALGRNTNYPHTITLRRONRYERLHNEEGTWIFDAKELEALAVQYQRLYSKEDVL
QVYNPLPDGSGELRLTRREVSLNKPLETEVEVAIRSMGKYKAPGPDGQPIFYQGW
EYVNSVIRFALDFFSSGILPPOTNDALVVJIPKVPKPERNQFRPISLGNALFKMIT
KMVLRKKVIGKLGISQSSFIQGRSLQNIIVVQEAHSLNKKKORKEKCHMIDR
AVAKKKSIGLSQGGPKISHTGFADDLIFAEASVQIRVIRITIEFCIASQKVS
LDKSKIPFSKNSRDLEKLTKSGSGIKSTRELKYLQMPLQRRINKDTFGYDERVS
SRLAGNKGRLSPFAGRLTLKSVLSLPIHTMSTISLPQSTLEGLKLAPVPLKSSA
EKKLHLVANDRVCLPKSEGGILGRTSKCNKALVSKVGRKLINDRYSLWRLRSKY
RVGLREVSGSRWVNGRDIIIPSDNWLSEALINRAVIEPNSKELRYKADKAN
GLGWKLPKIPY-SYHTRIEFAAVVDSTGARDLSWGSADGQVFTVKSAYKLTED
HDPRLPNNAEPDRIRVVALERYKFTLWHTGDTSVQCKGQDETILHVKKCPKSLAG
IMRRIQVQVSQYDFNCSIFPGMLVYNLGMKNATGYAWATLFAIVVMKMKQCYVPE
GEVKKCRDRYKFTKGLAAEYSHAHATHSQNGRLTRVRLVANKPPDGEWVKLNIDGA
SRGNLGLATTGGRDGLGHCWGGFALDQVCNPLAEJMGVYVGLYMAHERRTIVE
LEVDSELVGLTGTGSDTHSLSLFLVKLGHGFL"
exon: 10891..11025
/gene="AT4g08630"
/number=1
intron: 11026..11243
/gene="AT4g08630"
/number=1
exon: 11244..11393
/gene="AT4g08630"
/number=2
intron: 11394..11425
/gene="AT4g08630"
/number=2
exon: 11430..12120

```

```

/gene="AT4g08830"
/number=3
intron: 12121..12360
/gene="AT4g08830"
/number=3
intron: 12361..13089
/gene="AT4g08830"
/number=4
intron: 13090..13161
/gene="AT4g08830"
/number=4
exon: 13162..13585
/gene="AT4g08830"
/number=5
intron: 13586..13633
/gene="AT4g08830"
/number=5
exon: 13634..14348
/gene="AT4g08830"
/number=6
repeat_region: 16002..16053
gene: 18436..120750
/gene="AT4g08840"
CDS: join(18436..19424,19870..20032,20124..20287,20492..20750)
/gene="AT4g08840"
/notes="similarity to probable regulatory protein Y3L014w
Saccharomyces cerevisiae, PIR2.S64016
Contains Prokaryotic membrane lipoprotein lipid attachment
site AA82-92"
/codon_start=1
/product="putative protein"
/protein_id="CA878009.1"
/db_xref="GI:7267527"
/translation="MDGFFPGDLQRGSFTDGLGFPPLTSSVNGHFHSGRTINP
FLNRRKLDITSLMADGVDMGLCONLSKMSIDERSNFNHSFSYGCYQGRESFHHG
EASSMRGFGVGDYHRRPQDLRVASFHESASMSYVGDSYHRLPFLAQFASNP
NPRCFETNSLLNDRYMLREHFNQOIRRDFSLVPQKSPAPHEERILPPFSAMGCSR
EDVNGKAKMKNKEDSLDLASMYDSYVYLMADGLGRLLQKLFYDEGNFVDMILFK
EYNNVIELGTDPEGNLIQKLIENCNEFORTQILRLTSKPLAVKLSINNYGIRVV
QKLIETVTRKEQISLVKVSALVFGFLSLFRELNGNHVILNCLFFSPDNKFLBEAKK
FCIEATIRHGCCVLRQSVSYSGEHEKLVDEISNLSLLAQDPFGNTLVGYIIIEKK
VGVNVYLFELRGNVYKVLATQKFGSHVVEKLRYPESRSQIVNELVSVLNFYGLQDP
VANYIQCLASKTIVS"
exon: 18436..19424
/gene="AT4g08840"
/number=1
intron: 19425..19869
/gene="AT4g08840"
/number=1
exon: 19870..20032
/gene="AT4g08840"
/number=2
intron: 20033..20123
/gene="AT4g08840"
/number=2
exon: 20124..20287
/gene="AT4g08840"
/number=3
intron: 20288..20491
/gene="AT4g08840"
/number=3
exon: 20492..20750
/gene="AT4g08840"
/number=4
intron: 23831..27580
/gene="AT4g08850"
complement(join(23831..24120,24787..27580))
/gene="AT4g08850"
complement(join(23831..24120,24787..27580))
/gene="AT4g08850"
/notes="Est GB:F1413 is reverse and marks the end of the
gene
similarity to receptor protein kinase-like protein -

```

Arabidopsis thaliana, P12.T05898
Contains Prokaryotic membrane lipid attachment
site AA:2-22:Protein kinases signatures and profile
AA763-784:Protein kinases signatures and profile
AA883-895
Contains EST db:FL4412, Z17737, R84191, T22574, AA394763,
AA395692, P4413*

Query Match 29.4% Score 499.8; DB 8; Length 179771;
Best Local Similarity 90.6%; Pred. No. 1.7e-77;
Matches 557; Conservative 0; Mismatches 52; Indels 6; Gaps 2;

```
QY 1 TGGGTTTATTCGATACATGACAAATATTATTTATTCATGAGTCTTTTATTCGATAG 60
Db 17646 TCGGTTTATTCGATACATGACAAATATTATTTATTCATGAGTCTTTTATTCGATAG 17597
QY 61 CATGACAAATATTAATATATATATCATGCTTAAATACATGCTTCTTAAATACATGCTT 120
Db 17586 CATGACAAATATTAATATATATATCATGCTTAAATACATGCTTCTTAAATACATGCTT 17527
QY 121 TTAATATCAGACATTTTCTTTTAAATCAATCTTATCTTTATATCAGACATGAC 160
Db 17526 TTAATATCAGACATTTTCTTTTAAATCAATCTTATCTTTATATCAGACATGAC 17457
QY 181 GGAAATTCAGGTAAAGAGAAATTAAGAAATGAGATAGAGATTTCTATCGAAAA 240
Db 17466 GGAAATTCAGGTAAAGAGAAATTAAGAAATGAGATAGAGATTTCTATCGAAAA 17407
QY 241 AGAAAGAGACATGATGATGACAAATAAAGAGATATGATGATATTTATGAGAG 300
Db 17406 ATAAAGAGGACATGATGATGACAAATAAAGAGATATGATGATATTTATGAGAG 17347
QY 301 GGGTCAAGATATTTTCTTTTAAAGAGGAGAGAAATAGAAATAGAAAGAAATGACATG 355
Db 17346 GGGTCAAGATATTTCTTTTAAAGAGGAGAGAAATAGAAATAGAAAGAAATGACATG 17287
QY 356 GTGAATCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 415
Db 17286 GTGAATCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 17227
QY 416 AGTCTCTGTAAGACATGAAAGAGAAATAGAAAGAGAGAGAGAGAGAGAGAGAGAG 475
Db 17226 AGTCTCTGTAAGACATGAAAGAGAAATAGAAAGAGAGAGAGAGAGAGAGAGAGAG 17167
QY 476 ATAGACTAATCTTTCGCAAAATCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
Db 17166 ATAGACTAATCTTTCGCAAAATCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 17106
QY 536 TTGCTGATCTTTTGAAGTCAGAAATGATTTCTTACATATCTTAAATATACCCGAT 595
Db 17107 TTGCTGATCTTTTGAAGTCAGAAATGATTTCTTACATATCTTAAATATACCCGAT 17045
QY 596 ACCAATTTTCCACA 610
Db 17047 ATAACGTTTGCAAA 17033
```

RESULT 8
AF071527 95576 bp DNA linear P12 06-MAY-1999
LOCUS Arabidopsis thaliana BAC F9H3, from chromosome IV near 18.8 cm,
DEFINITION complete sequence.
ACCESSION AF071527
VERSION AF071527.1 GI:3236479
KEYWORDS HTG.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 95519)
Huang,E.N., Parnell,L.D., de la Bastide,M., Schutz,K.,
Habermann,K., Dedhia,N.N. and McCombie,W.R.

Genomic sequence of Arabidopsis thaliana BAC F9H3, chromosome IV,
18.8 cm
Unpublished
2 (bases 1 to 95519)
Huang,E.N., Parnell,L.D., de la Bastide,M., Schutz,K.,
Habermann,K., Dedhia,N.N. and McCombie,W.R.
Direct Submission
Submitted (10-JUN-1998) Cold Spring Harbor Laboratories, Lita
Annenberg Hazen Genome Center, 1, Bungtown Road, Cold Spring
Harbor, NY 11724, USA
3 (bases 1 to 95519)
Parnell,L.D. and McCombie,W.R.
Direct Submission
Submitted (31-DEC-1998) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
Arabidopsis thaliana BAC F9H3 from chromosome 4, near 18.8 cm
4 (bases 1 to 95519)
Parnell,L.D. and McCombie,W.R.
Direct Submission
Submitted (01-FEB-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
Arabidopsis thaliana BAC F9H3 from chromosome IV near 18.5 cm
5 (bases 1 to 95519)
Parnell,L.D.
Direct Submission
Submitted (06-MAY-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724
Arabidopsis thaliana BAC F9H3 from chromosome IV near 18.8 cm
BAC F9H3 is assigned between YACs C14A7 and C1C8B1 and maps to
near 18.8 cm on the Lister & Dean RI map. Position 1 of F9H3 is
oriented toward the centromere and position 95519 is oriented
toward the telomere. For more information on the mapping,
sequencing and annotation of F9H3, please see
<http://www.cshl.org/arabweb/F9H3-titlepage.html>. A graphic view of
our annotation is also available at this url. Gene models are built
with exons predicted by GENSCAN
(<http://CCR-061.mit.edu/GENSCAN.html>), MZEF
(<http://www.cshl.org/genefinder>) and GRAT,
(<http://compbio.ornl.gov/tools/index.shtml>) and with splice sites
predicted by NetPlantGene
(<http://www.cbs.dtu.dk/netgene/cbsnetpgene.html>). Alternate exons
not used in building the gene models are presented on the web pages
associated with F9H3. Genes are numbered according to the scheme
BAC.gene.number. Typically, these numbers progress from 1 upwards
as one moves from position 1 of the BAC. Protein sequences encoded
by the genes are assigned to a functional category with the aid of
similarity searches and comparison to the Prosite
(<http://expasy.hcuge.ch/sprot/prosite.html>) and Pfam
(<http://pfam.wustl.edu/>) libraries. A description of these
categories can be found at
<http://antjac.mips.biochem.mpg.de/arabi/>. Genomic repeats are
typically located by TBLASTX analysis and an attempt is made to
classify the function of each repeat as either transposon, putative
microsatellite, LINE, direct repeat, centromeric repeat, etc.

If you have any questions or contradictory or contradictory evidence
concerning the annotation of F9H3, please direct email to Larry
Parnell at parnell@cshl.org.
BAC F9H3 was sequenced as part of the Arabidopsis genome sequencing
effort of the Cold Spring Harbor Consortium. For
additional information, please see <http://www.cshl.org/arabweb>.
Fingerprint data indicate F9H3 overlaps with F4C21 toward the
telomere and with T5123 toward the centromere. F9H3 contains marker
T1233 at 18.8 cm on the Lister & Dean RI map.

1. 95519
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"

FEATURES
SOURCE

```

/misc_feature
1..31783
/translation="MSLKHRTVPQTPGRPLAAEAAGIEEPYNIIPVNNLLADHPSLR
FQVRAAAALKTVDLRPPYVWMSHYDLDLMLALFTGCFKQDNVNRNREHVMVHLA
NAQRLSPDDNIDSLSAVRRFRKRLNANYSWCSYLGKKNINISNRPDSRREL
JYGLVLIWGEAANLRFWPECIFYIFHNMANSELNKILDCDENTGQYPLPSLGEN
AFLTVGKPIYOTQAEIDESKNGYVHCKWRNYDDINEYFWTDFGSKLKNPLDGS
NFKSGKSGVGTGQVRRRTFFLYRSFORLWMLALFLQAAIIVAFSEKPKOTSSVTR
OLNWLAKRDOVRLTLVFLTWGSMGLLOAVIDAASQYPLVSRRETKHFRMLMKVIA
AAMVIAFTVLTINWKKROHNSNAATTKIYQFYAYAGFLVPELALALFLIPW
MRNLEETNMKIFPALTMPFGKSFVRGRLGVNDNIKISTFWILAKITFSYFL
QVPMIKPSKLWNLKDVDEWHQFYGDSNSVAIPLPVLVLIQWIAIYSS
IVGAVVGLFDHLGEIDMGOLRLRFQFAAIQNLPEEQLNARGFNKFGKDG1HR
LKRYVGRPRFKKLESNOVEANKFALINWEIIILAFREEDIVSDREVELLEPKNSMDV
TVIRWPCFELICNELLALSOARELIDAPKWLWHKICKNEYRCAVCAVADS1KHLL
SI1KVDTREHSIIVFQIINOSIQSEOPTKTFVDPLLPKIYETIOLKLVGNDESD
SGRVNVLQSLIETIATROFFIEKTTQESNEGTPROPASKLLFONALRCPDASNEC
FYPRVRLTLTSRDSMSVNVNLEARRIAFFSNLFMNMHPAPOVEMKAFSVIT
FYSEEVYSKEQLRNETEDG-STLYILOTIYADENKFKERMHREGIKTDSLELWTK
LRDLKASVGTOTLARTVGMVYVYALKMLAFLOSASEMDIREGAQELGVSVMQOG
ELGGQSDGVSENDSSLRSSSVSTLYKGHEYGALMKFTYVACQIVGSOAKKE
POAEELIYLMKQNEALRIAYVDEVPAGRGETDIYVLYKYDHOLEVEYELFVKLPGP
VKLGEKPNQNHAMITRGDAVOTIDMODSYFEALAMNLLQYENIHYGIRKETI
LVGREHIFTGSVSSLAWSAQETSEVTLGQRYLANPLKVRMHYHPDVFQWFLSR
GGISKASRVINISIDITAFGNCFLRGNTVHEYIQVGRDGLNQLISFNAKVASG
NGEQLSRDVIYRLGRLDFPRLSFYTVGFPTMTWILTYYAFGLGRVYLALSGV
EKSALADSTDNAALGVILNQOPILOGLFTALPMIVENSLEEGFLATMNEIRMOIQ
LSAVFYTFSGMTRAHYFGRTILHGKAKYRATGRGVVHEHGTENTYRLARSHFVKAI
LGNLIVYASHPSIARSDLIYTAMTISNFWLSWIMAPFVFNPSGFWMLKTYDDE
DFNMWYOGRISTKSEQSWKMWYEQDHLRNTKAGLFEVILVLRFFQYQIVY
OLKANGSLFYLFWSWIFYFAIFVLVYQIYARDKYSAKAHIRYLVOLFVLIIVAI
LV1VALLETFHFSDIFDITSLAIFTGIGILLIAOTQRKWKNTYFNNVYVSVARM
YDILGGLIMVPVAFELSWMPGFGSMOTRILFNEAFSRGLRIMQIVTGKSKGDV"
complement(20861..22302)
/notes="F9H3.17"
/feature="F9H3.17"
/notes="encodes hypothetical protein: gene model last
edited on 5 Jan 99"
/evidence=not_experimental
complement(join(20861..21094,21937..22035,22145..22302));
/genes="F9H3.17"
/notes="similar to F21B7.5, GenBank accession number
AC002560; functional catalog ID=99"

CDS
29.2%; Score 497; DB 8; Length 95519;
Best Local Similarity 92.5%; Pred. No. 5,9e-77;
Matches 545; Conservative 0; Mismatches 40; Indels 4; Gaps 2;

Cy 1 TGGGTTTATTGGATACATGACAAATATTTATTATTCATGAGTTTATTATGGAGTAG 60
Db 23303 TCAGTTTATTGGATACATGACAAATATTTATATATACATGAATTTTGTGGAGTAG 23362

Cy 61 CATGACAAATATTAATATATCAGTGTGTAATAACATGTTTGTTCITAAATACATGCAAT 120
Db 23363 CATGACAAATATTAATATATCAGTGTGTAATAACATGTTTGTTCITAAATACATGCAAT 23422

Cy 121 TAAAAACAGACATTTGTTTAAAAACAAATCTTAATCTCTTAATACACAGCATGAC 180
Db 23423 TAAAAACAGACATTTGTTTAAAAACAAATATAATCTCTTAATACACAGCATGAT 23482

Cy 161 GAAAAATCAGGTAAAAGAGAAATTAAGNATGAGATAGAGAGATTTCTATGGAAAA 240
Db 23483 GAAAAATCAGGTAAAAGAGAAATTAAGNATGAGATAGAGAGATTTCTATGGAAAA 23542

Cy 241 AGAAGAGAGACATGTAGGTGAACAAAATAAAGAGATATCATGATATATTTATGAGAG 300
Db 23543 AGAAGAGAGACATGTAGGTGAACAAAATAAAGAGATATCATGATATTTATGAGAG 23602

Cy 301 GTGIGTAGGATTTATTTAGGAGAGGAGAGAAATAGAAAAGAAAATGACATGGTCAA 360
Db 23603 GTGIGTAGGAGATTTATTTAGGAGAGGAGAGAAATAGAAAAGAAAATGACATGGTCAA 23662;

Cy 361 TCTGAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAGAGAACTCCATCGGTAAAGTC 420
Db 23662 TTTCAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAGAGAACTTCATGGCTAAAGTT 23721

```

```

/chromosome="iv"
/map="18.8 cM"
/cloc="BAC F9H3"
/notes="hybridizing YAC is CUC8B.1"
1..31783
/notes="overlap with 75L23, GenBank accession number:
AC005142, from position 26 to 31806"
851..1719
/genes="F9H3.20"
/pseudo
/evidence=not_experimental
complement(981..1120)
/notes="function=putative glucan synthase; similar to F6K23, GenBank
accession number AF058919"
/rpt_type=dispersed
3173..3969
/notes="function=putative calcium channel; genomic copy of
EST W43766; genomic copy of EST T43251; EST T43251
apparently encodes intronic sequence; possibly alternate
exon 1 from position 12772 to 12845 is predicted by a
NetPlantGene splice donor but has no EST coverage; gene
model last edited on 27 Jan 99"
/evidence=not_experimental
complement(join(6553..5746,6833..5870,6991..7062,
7175..7261,7436..7526,7682..7745,7985..8153,8229..8322,
8380..8487,8757..8814,8891..9034,9127..9192,9311..9383,
9530..9591,9745..9822,9929..10034,10127..10293,
10331..10451,10557..10580,10644..10774,10850..10930,
11292..11403,11500..11643,11972..12091))
/genes="F9H3.19"
/notes="functional catalog ID=07.01"
/codon_start=1
/evidence=not_experimental
/product="putative calcium channel"
/protein_id="AADI1598.1"
/db_xref="GI:4206210"
/translation="MEPLHGRDGLGGGTRVRSSEATHGTFQKAAALVLUAFEG
IGLVEILDQSSGESARYIFITRLDLINSNYFALLFLNFFEQPLWCEKNPKPSCK
DRYYLDELPLYNAESIIEASRIFTWRSJNLVKVAVYILFVJVLVDFLY1.SPL
AFDELPLRIAPYRVIIFILSIRQCADHVMELRDLVLSGLMIGTYLNLALWMLFL
LFASWIAFVPEDTQOGLVFTSYGATLYQMFILFTSNPDVWIPAYKLSRMSVFP
VLYVLIGYFVTLILAVVYDSFEOLAKOVSGNDQKRMLEKAFGLIPSDKNGETD
KNQCIKIFEOLTYRTPKISKKEFGTFDELDTROFKINKEFAALCQALALPFOK
EVSLEFHPQIHTHLSQOLRAFSPNFGYAIISFILINFVAVVETIDIESS
AQKWQVAPVFGWYIILEMALKITYGFNWREGANRDFLVWIV1.GEATFIT
PDNFTFTSGEMWYIILLRLIMVORYRAFIAFIT1.PSIMPYLGSTIFGV
LCVCS1.GVQVQGLVGNAGKKLETELAEDDYLLENFNPNMGWIT1.FNLLVNGNKQ
VMMESYKDLTGKWSITVYFVSFYVIT1.LILLNLVAVFVLEAFET1.ELDEHEKQCOQO
SOEKNRNRSGKSRQORVDTLLHMLQDELSPKSTSD1"
complement(13149..18831)
/genes="F9H3.18"
/notes="encodes putative glucan synthase component: gene
model last edited on 5 Jan 99"
/evidence=not_experimental
complement(join(13149..14597,14778..16800,16961..18831))
/genes="F9H3.18"
/notes="similar to 1,3-beta glucan synthase; functional
catalog ID=01.05.99; functional catalog ID=09.01"
/codon_start=1
/evidence=not_experimental
/product="putative glucan synthase component"
/protein_id="AADI1597.1"
/db_xref="GI:4206209"

```



```

/Note="function terminal_flanking_direct_repeat: cognate B
of flanking direct repeat of 18r transposon from position
5:45 to 6873; cognate A resides from position 5:45 to
5:49"
repeat_region
/rpt_type=flanking
/complement(7885..7924)
/Note="function:putative_microsatellite"
/rpt_type=tandem
/rpt_unit=AT
/complement(7905..8519)
/Note="function-unclassified; similar to MX110. GenBank
accession number AB005248"
repeat_region
/rpt_type=dispersed
/9482..10923"
/gene="TSL23.3"
/Note="encodes hypothetical protein; identical to F9H3.17,
GenBank accession number AF071527; gene model last edited
on 5 Jan 99"
CDS
/evidence=not_experimental;
/Join(9482..9639,9745..9847,10690..10923)
/gene="TSL23.3"
/Note="identical to F9H3.17, GenBank accession number
AF071527; similar to F2187.5, GenBank accession number
AC002560; functional catalog ID=99"
/codon_start=1
/evidence=not_experimental;
/product="hypothetical protein"
/protein_id="AAB15310.1"
/db_xref="GI:4263041"
/translation="MVKLTIKRGGVLRLAAGAAALIVMTSRERASFLA;SLEA
KYTDMAAKFYVIANAVSVSYFLVFLPKESLWKEVVDLDMTLMELTSS;GAALA
VAQVCKGNANAGWPIGQVPKFCQITGAL;AGEVALVLYVLLLSLRAVDPTL
LQAS"
/gene="TSL23.3"
/gene="TSL23.4"
/Note="encodes putative glucan synthase component;
identical to F9H3.18, GenBank accession number AF071527;
gene model last edited on 5 Jan 99"
/evidence=not_experimental
/Join(12953..14823,14964..17006,17187..18635)
/gene="TSL23.4"
/Note="similar to 1,3-beta glucan synthase; identical to
F9H3.18, GenBank accession number AF071527; functional
catalog ID=01.05.99; functional catalog ID=05.01"
/codon_start=1
/evidence=not_experimental
/product="putative glucan synthase component"
/protein_id="AAB15311.1"
/db_xref="GI:4263042"
/trans-ation="MSLRHTVPPQIGRLPAEAVSIEEPYNI;PVANLLADHPSIR
FEVPAARALXIVGLREPPVQVRSHVYDCLMLALPFGOKQNVKRCFEHNV;HIA
NAQMLSPPDN;DLSNVAVRFRKLLANYSKCSVLGKSK;NISDRNDSREPL
LVGLFLLWGAALNRMPEDICITFHNASSELKRL;ECUDENIGQYPLPS;SGEN
AFLTGVPKPIYDIOAEIDESKNGIVAHCKWKNYDINEYFW;DRCPKSKWPDLSG
NFFKSRKSGVGVITREFTFYLYSRFDRLMIALQLQAAIIVAMEKPKDLSVY
QJWNAK;KAGQVOVRLITVLTWSKRLQAVLDAASQYPLVSKR;KRIFFRMLKRV
AAWVAIVTNYLTNWKOKROGRKSNATTKIYQFLAVAGFLVLAALALF;TPW
KRNPLEETNKKIFFALTWPQKSEVGRGLREGLYDNKYSTFIYFLATKFT;SYFL
QKMPKTLKLLWNLDLQAEVQFQDGNRSFVAL;MLPVYLLMDIQ;WFA;YSS
IVGAVGFDFHLGE;RMQCC;RLRFOFFASQ;FNLMPEDQLNARGFNKFD;IHR
LKLYRGFGPKKLESQVNAKFKALIMNEILAFREEDIVSDREVELLE;LPKNSWDY
TVIRNPFCLLNELLALSOARELIDAPQKMLWHK;ICKNYRRCVAVVAYS;IKILL
SIUKVTEBSHSLTYFOILOSIOSEQFTKTFVRVDLIPKIYETLCKLVG;VNDRETD
SGRVNVLQSLVEIATQRFIEKKTEQLSNEGLTPRDPASKLLQFNALR;LDASNE
FYQVRRLTILTSRDSHSPVNDERRRIAFFNSLPMNPHAPQVQKMAF;SVIT
PYISEVYYSKQILNETEDGSTLYLQTIYADENKFKRMHREGIKIDSELWTIK
LRDLRLWAGYRGTLARTVGRMMYVYRLAKMLAF;LDSASEMDIRSGAOLGSGVRLQ
ELGQSDGFGVSENDSSLRSSSVSYLYKGHEGYTALMKFTYVAVACQYLGSKAKKE
PQAEIILMKQNEALRIAYVDEVPAGRGETDYYSVLVYDHOLESKEVEIFRVKLPGP

```

Query Match

Best Local Similarity 92.5%; Pred. No. 5.6e-77;

Matches 545; Conservative 0; Mismatches 40; Indels 4; Gaps 2;

```

QY 1 TGGGTTTATTCGATACATACGACAAATATTATTATTCATCAGTTTATTTCGATAG 60
DB 948 TGAATTTTATTCGATACATACGACAAATATTATTATTCATCAGTTTATTTCGATAG 8422
QY 57 CATGACAAATATTAAATATATATATATATATATATATATATATATATATATATAT 120
DB 842 CATGACAAATATTAAATATATATATATATATATATATATATATATATATATATAT 8362
QY 121 TTAATAACAGACATTTTGTAAATCAATCTTAATCTCTTATATATCAACAGCATGAC 180
DB 8361 TTAATAACAGACATTTTGTAAATCAATCTTAATCTCTTATATCAACAGCATGAT 8352
QY 193 GGAATATTCAGGTAAACAGACAAATTAAGAAATGAGATAGAGAGATTTCTA;GGAAA 240
DB 8311 GAAATATTCAGGTAAACAGACAAATTAAGAAATGAGATAGAGAGATTTCTATGAAA 8242
QY 241 AGAAGAGAGACATGTAGGTGAACAAAATAAGAGATATGATGATATATTTTAIGAGAG 300
DB 8241 ACAAGAGAGACATGTGGTGAACAAAATAAGAGATATGATGATATTTTAIGAGAG 8182
QY 331 GTGGTGAAGATATTATTTAGAGAGGGAGAGAGAAATAGAAAAGAAATGACATGTTGAA 360
DB 8131 GTGGTGAAGATATTATTCAGGAGGGAGAGAAATAGAAAAGAAATGACATGTTGAA 8123
QY 361 TCTGAGAGAGATGAATTTGTCTTAAGATGAAGAGAGAGAGAACTCATGCTTAAAGTC 420
DB 8122 TTGAGAGAGATGAATTTGTCTTAAGATGAAGAGAGAGAGAACTCATGCTTAAAGTC 8063
QY 421 TCGTAAACAGATGAAAAGAGAAACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 8062 TCGTAAACAGATGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8006
QY 481 CTACATATGCGCAAAATTTCTGTAGCGCGCAAAATCTACTTCTTACATATCTCTAAAATATA 540
DB 8065 CTAAATATATGCGCAAAATTTCTGTAGCGCGCAAAATCTACTTCTTACATATCTCTAAAATATA 7946
QY 541 TATTCCTTTGAGTCAAAAGTTATCTTCTTACATATCTCTAAAATATA 589
DB 7945 TATTCCTTTGAGTCAAAAGTTATCTTCTTACATATCTCTAAAATATA 7897

ATCHKV9 159629 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 9.
ACCESSION AL151497
VERSION AL151497.2 GI:7270670
KEYWORDS Arabidopsis thaliana.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids 1; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 53060)
AUTHORS Spege, L.A., Huang, E.N., Nascimento, L.U., de la Bastide, M.,
Vil, D.M., Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A.,
Rodriguez, M., Shekher, M., Schutz, K., See, L.H., Swaby, I.,
Hadermann, K., Dedhia, N.N., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
Unpublished
JOURNAL 2 (bases 21275 to 137726)
REFERENCE 3 (bases 137729 to 137730)
AUTHORS Lamar, B., Stoneking, T., Stumpf, J., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
Unpublished
JOURNAL 4 (bases 1 to 159629)
REFERENCE 5 Arabidopsis sequencing project.
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:

```

lcmckee@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@jbsrc.ac.uk
 information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/prc/thal/>
 this fragment has an overlap with ATCHR1V8 at the 5' end and an
 overlap with ATCHR1V10 at the 3' end.

FEATURES

Source

1. 159529

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

6741..9588

/gene="AT4g03480"

complement(join(6741..7064,7161..7708,7783..8101,

8409..8866,9283..9343,9445..9588);)

/gene="AT4g03480"

complement(join(6741..7064,7161..7708,7783..8101,

8409..8866,9283..9343,9445..9588);)

/gene="AT4g03480"

/note="contains similarity to mammalian ankyrins

similarity to similar to A. thaliana hypothetical protein

F21B7.8, GenBank accession number AC002560

Contains ATP/GTP-binding site motif A (P-loop) AA297-304

contains EST gb:37081, 742092"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAB77834.1"

/db_xref="GI:7270671"

/translation="MYRSHLRATPEASPLKPASSPNCRLSLSPVELRVPFIRDEBP
 KYVDIFKSYONSPELPCCDKSPCLSLRESLEPLEGVEYQWNPPIFSMRAQNI
 ELLEKLKSVETMAGLKSDGGSVHLAASGHELVNITECHCLLEPNSKTCQIP
 LHVAAAGSAVVKALVASVYFSRVPPEEDRLNIYVKQIDGETPLHAAIKGIRE
 KAEERIKLSLSLHMHWRSCISFSDASTQMETAAALVNADQHASFLANKDQSP
 IYLVADAGNLSVLRAVLRNPGNKIQGKTSLASQLEGRKSLLEALKAKKATDVAVIL
 NDPLSVNERDEGRKTSVSGAMGYKIGICKLRSTKSVYCEPKDGSFPIHMAVEK
 GHLKVKVETILKRCPSKELVNGQNMGLIAAKSAKVSFLLSYIRLLEENMLLEEQ
 DYDGNAPLHLATINRCRTVDLAAPASTETKILNIQKDGURPDIAEJNLQPDYVL
 RERLTMLVLCVYAKSKGWLPTSGMTIHSRSEPLDARKYKDIKAILILVATLVATV
 FAAGTIPCGFNSSAPNMGMATLADDSLTFFTVLDTLMAQSSVAIVAL."

complement(6741..7064)

/gene="AT4g03480"

/number=1

complement(7065..7160)

/gene="AT4g03480"

/number=1

complement(7161..7708)

/gene="AT4g03480"

/number=2

complement(7709..7782)

/gene="AT4g03480"

/number=2

complement(7783..8101)

/gene="AT4g03480"

/number=3

complement(8102..8408)

/gene="AT4g03480"

/number=3

complement(8409..8866)

/gene="AT4g03480"

/number=4

complement(8867..9282)

/gene="AT4g03480"

/number=4

complement(9283..9343)

/gene="AT4g03480"

/number=5

complement(9344..9444)

/gene="AT4g03480"

/number=5

complement(9445..9588)
 /gene="AT4g03480"
 /number=6
 :0335..13501
 /gene="AT4g03490"
 complement(join(10335..10525,10983..11078,11590..12133,
 12345..12500,12627..12772,12871..13501))
 /gene="AT4g03490"
 complement(join(10335..10525,10983..11078,11590..12133,
 12345..12500,12627..12772,12871..13501))
 /gene="AT4g03490"
 /note="contains similarity to mammalian ankyrins
 similarity to similar to A. thaliana hypothetical protein
 F21B7.8, GenBank accession number AC002560"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAB77834.1"
 /db_xref="GI:7270672"
 /translation="MGKVOEFKVMENEIPVLDOVIFQGNLILHLAIIYCHDHLVEE
 ILAYELAILERNKRLNCFVPSFSHYQTLIVRRNYKGDALHLVAAAGHKLIVGL--
 DCLQLQPDITIMVIGSEQMVFIGNIFRVSNNDGNTALHLKGNHVSLSQLREDSI
 CFLDKEDVSPLYMAAEGYVSVLVEHMLRGLDASFVGSVLCAAVKSCLNDILTAVLE
 SDSLVSERDSDGRTPLATASIGYDGVQHMLTFASSTQGNVLHVAASKGNARAV
 GYLLKSDVKRLINEQDIEGNTPLHLASSNSHPKVLIMMALVAAGTIRAPRVHLRAD
 YPLGTTDEDLILKTHKDRVNTLLVAVTLVATMAFAAGLVPLGYNTEFSKNVKSVE
 ESATFAYICNSIAVYAVISTVALIGTQLADLACMLITTFKFIPELLGFSIIAMSLAF
 VAGIYLVGHYKLAIFVLASGGFYLMALLLIIPYASPTFTILSRSLNSLYQNSKE
 DVDSVNLQVPAPTEELALIKAIQVNRCTQGNIGVTSLSYLLKHLKRVAFINLLILKLN
 HYSLVFSGHYDGPCKTYSQL"
 complement(10335..10525)
 /gene="AT4g03490"
 /number=1
 complement(10526..10982)
 /gene="AT4g03490"
 /number=1
 complement(10983..11078)
 /gene="AT4g03490"
 /number=2
 complement(11079..11589)
 /gene="AT4g03490"
 /number=2
 complement(11590..12133)
 /gene="AT4g03490"
 /number=3
 complement(12134..12344)
 /gene="AT4g03490"
 /number=3
 complement(12345..12500)
 /gene="AT4g03490"
 /number=4
 complement(12501..12626)
 /gene="AT4g03490"
 /number=4
 complement(12627..12772)
 /gene="AT4g03490"
 /number=5
 complement(12773..12870)
 /gene="AT4g03490"
 /number=5
 complement(12871..13501)
 /gene="AT4g03490"
 /number=6
 13514..15185
 /note="marker mi233"
 14170..17288
 /gene="AT4g03500"
 join(14170..14204,15103..15848,15939..16477,16550..17283)
 /gene="AT4g03500"
 /note="hypothetical protein with some similarity to
 ankyrin
 similarity to"
 /codon_start=1
 /product="hypothetical protein"

/protein_id "CAB77835.1"
 /gi_xref "gi:7270673"
 /translation "MKNESRTSEEISKGRIFSGVGFSKGRPDVHLDERRVAINHQAIPK
 NVNRTSPSLDSTLFDSTPTKDKDPKIMAAVAGKNVLRSNNSYISVAPLIVNDRG
 NTLHLAASSGHVSLVRYIITQKCPGILLKSNMGMGEVALHLAAEAGHLDPVWNLIDPIN
 DISCTNPVAKRKYFAKKNQNDTALHVALKGRHEVASYLVLSAARKSPVANFGGSPF
 LLAIEAGHTSLVNTMCTGNELSSKVGGRSIVHAAKANKKDIIDALLSKDASLILN
 RDEGRTSSEFAGSGIYGYGFSYLLFDKNRKYVSYSDGILFPTHMAAKYGHVQIIEEL
 KHCPEATELDRQGNILHLAAKYGLKVIKFLSCCKDKNKKLLINEQVNCNNTNLH
 LATLNMHPKVVYNTDHWDLKKRYVIGTALDVAEENIDSSYVHOBLTWMAJLNA
 CAPASSTPTENLSPKPKDGGRYKORVWFLMLVATLVATMTTGAFTLPGGVNLSFP
 HLGNAVLAARATAGVFLVCDTILAMYSITIVAILMAWQGLSLIKANINALEPFLICL
 ALGSMSTAFMAGTYVAVSHPTLLGYFVLGIGITFLVLELLLVFVYSPYAHAGPFLIRH
 IFYFPYFLKLLAAGDNKNINVDVYTAStb"
 14170..14204
 /gene="AT4q03500"
 /number=1
 14205..15102
 /gene="AT4q03500"
 /number=1
 15103..15848
 /gene="AT4q03500"
 /number=2
 15849..15938
 /gene="AT4q03500"
 exon
 intron
 exon
 intron

```

Query Match      23.2%; Score 497; DB 8; Length 159629;
Best Local Similarity 92.5%; Pred. No. 5.3e-77;
Matches 545; Conservative 0; Mismatches 40; Indels 4; Gaps 2;

1 TGGGTTTTTATGTGAACATGACAAATATTATTATTTATTCATCAGTTTTATTCGGATAG 60
25758 TGAGTTTTTATGTGNTACATGACAAATATTATATATACATGANTTTTGTTGGATAG 25659

61 CATGACAATATATATATATCATGTGTTAAATACATGTTTCTGCTTTAAATACATGCAAT 120
25658 CATGACAATATATATATATCATGTGTTAAATACATGTTTCTGCTTTAAATACATGCAAT 25639

121 TAAAAATCAGACATTTGGTTTTAAAAATCAAAATCTTAATCTTTATATCAACAGCAATGAC 180
25638 TAAAAATCAGACATTTGGTTTTAAAAATCAAAATATATCTCTTATATCAACAGCAATGAT 25579

181 GGAAATTCAGGTTAAAAAGAGAAAATAAAGAAATGAGAGATAGACAGATTTCTATCGSAAA 240
25578 GAAGAAATTCAGGCAAAAAGAGAAAACAAGAAATGAAGAGATAGACAGATTTCTATGGAAA 25519

241 AGAAAGACAGACACATGTAGGTGACAAATTAAGAGATATCATGATATATTCTATGACAG 300
25518 AGAAAGACAGACATGTGGGTGACAAATTAAGAGATATCATGATATATTCTATGACAG 25459

301 GTGTGAAGATTATTTTAGSAGGGGAGAGAAATAGAAAAAGAAAAATGACATGGTGA 360
25458 GTGTGGAAGA-TATCTTAGSAGGGGAGAGAAAGTAGAAAAAGAAAAATGACATGGTGA 25400

361 TCTGAAGAGATGAATTCGTGTTAAAGATGAAGAGAGAGAGAACTCATGSCCTAAATTC 420
25399 TTTGAAGAGATGAAATTCGTGTTAAAGATGAAGAGAGAGAGAACTTCATGSCCTAAAGAT 25340

421 TCGTAAAGAAGATGAAAAAGAAACAAAAGAGAGAGAGAGAAAGAGAAAGGCATAAATAGA 480
25339 TCGTAAACAACATGAAAAAAGAAAAAGAA-----AAAGAGAAAAAGAGAAAGAAATAAATAGA 25283

481 CTAACTATTTGCCAAAATTTCTGTACCCGCAAAATTAATTTTGGTCCAAAGGTTATTTTGTG 540
25282 CTAAATATTTGCCAAAATTTCTGTACGACAGCAAGTACTATTGGTCCCAAGTTA-TTTGTT 25223

541 TATTCCTTTTGAAGTCAAAAGTTTATTTCTTACATATACTCTCAAAAATATA 589
25222 TATTCCTTTTCAAGTCAAAAGTTATTTTGTGATATACTCTATATATATA 25174

RESULT 11
CH 25250185

```

CSULT 11
CH250185

LOCUS	369 bp	mrna	linear	PLN 15-AUG-2000
DEFINITION	Arabidopsis thaliana mRNA for NIMIN-2 protein (nimin-2 gene).			
ACCESSION	AJ250185			
VERSION	AJ250185.1 GI:12057155			
KEYWORDS	nimin-2 gene; NIMIN-2 protein.			
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
AUTHORS	1. (bases 1 to 369)			
TITLE	Weigelt R.R., Rauscher, C., Pfitzner, A.J. and Pfitzner, U.M.			
JOURNAL	NIMIN-1, NIMIN-2 and NIMIN-3, members of a novel family of proteins from Arabidopsis that interact with NPRI/NIMI, a key regulator of systemic acquired resistance in plants			
MEDLINE	Plant Mol. Biol. 46 (2), 143-160 (2001)			
FORMED	2135044			
REFERENCE	1142051			
AUTHORS	2 (bases 1 to 369)			
TITLE	Pfitzner, U.M.			
JOURNAL	Direct Submission			
COMMENT:	Submitted (07-OCT-1999) Pfitzner U.M., General Virology, Institute of Genetics, Emil Wolff Str. 14, 70599 Stuttgart, GERMANY			
FEATURES	Related sequence: AB023041.			
source	Location/Qualifiers			
gene	1..369			
CDS	/organism="Arabidopsis thaliana"			
	/cultivar="Col-0"			
	/db_xref="taxon:3702"			
	/tissue_type="green vegetative tissue"			
	/dev_stage="3-wk-old"			
	1..369			
	/gene="nimin-2"			
	1..369			
	/gene="nimin-2"			
	/function="putative role in SAR"			
	/codon_start=1			
	/evidence="experimental"			
	/product="NIMIN-2 protein"			
	/protein_id="CAC19845.1"			
	/db_xref="GI:12057156"			
	/db_xref="SPTREMBL:Q9LUA3"			
	/translation="MNNSLKKEEVEEDNKGKSPSTEVYRTVIEEVDFFKILRRVHVATRTAKVNGVAEGELPSKKRKSQNLGLRNSLDCNGVADGEFDEINRVL			
BASE COUNT	113 a	53 c	121 g	82 t
ORIGIN				
Query Match:	21.7%; Score 369; DB 8; Length 369;			
Best local similarity	100.0%; Pred. No. 2.5e-54;			
Matches 359; Conservative	0;	Mismatches	0;	Indels
	0;	Gaps	0;	
QY	853	ATCAACAAC	CTTTGAGGAAGAGACGGGTAGACAAGATACGGAATCTGACGGT	922
DB	1	ATGAACAAC	CTTTGAGGAAGAGAACGGGTAGACAAGATACGGAATCTGACGGT	60
QY	923	AACAGAGGGAAC	CGGTGACGGAAGTTGTTCCGACGGTAACGGAAGAGGTGGATGAG	982
DB	51	AACAGAGGGAAC	CGGTGACGGAAGTTGTTCCGACGGTAACGGAAGAGGTGGATGAG	120
QY	983	TTTTTCAGATAT	TACGGAGGTACACGTGGGACACGACGGTTCGGAAGTTAACGGC	1042
DB	121	TTTTTCAGATAT	TACGGAGGTACACGTGGGACACGACGGTTCGGAAGTTAACGGC	180
QY	1043	GGTGTGCTG	TGAGGAGGTTACCGTCTAACAAGAGGAACGGAGTCAGATCTGGGTG	1102
DB	131	GGTGTGCTG	TGAGGAGGTTACCGTCTAACAAGAGGAACGGAGTCAGATCTGGGTG	240
QY	1103	AGAAATCGT	TGATTCACCGGTTTCGACGGGAAATTCGATGAGATTAATCGGGTC	1162
DB	241	AGAAATCGT	TGATTCACCGGTTTCGACGGGAAATTCGATGAGATTAATCGGGTC	300
QY	1163	GGGTTACAGG	GTGTTGGATCTCAACTGTAAACCGGAACACGACGCTTAGTITA	1222


```
/note="Similar to retrotransposon polypeptide"
/codon_start=1
/db_xref="GI:5080762"
/translacion="MAYNAGMGKSDRRKRFKKKEEKEDSDGERDFVGNFERICEAKA
CHEKLENDLRFVSTSLERTEIRAGKASEOPHDFQPKYKSSSIHSRRAPMESE
RYLNLPLDNCVSLAKIEMVLMGLILNGFOLWRFEFRVGRYHNDUKNLVLSL
TEPEDLSQVGLDVSHLEKIFVNGLRPEMKENIKKESRGLPNKVAVLRMESSAY
QMSRSGKNVARTINKSTKFPFLICSGAPVAHRKALKADNNLEWELGNSVSH
CSGCKGLVWLGVEGGMFMSLEGLVDVIGEWLELAVPKVLEIHESDLTGS
LSLSLAIMELPVVEVLSEFPFIIDHLEPFRGLEHSISLKPVSS:VRSYRY
PHATVLEKVMVENVITGPIPIVIELLDELHGM:FSKIDILCSGVH
QIRKPDILKTAFTLEGVVERFVMSGLTNAPATFOSSMNRISKLLKFLVLFVD
HILYSPVSHVHKH:REVLLQADNLKLYENMKKCSFGVGVX:GHIISAGOVVTS
IKTAMENKQVSNELKRLFGELGTGYIKFVRGYSIAKTLBELAKKAFJMLNQC
DAFQVQKAM:LA:VLGFPIDFVFEDTDFGTVGVGAVLMQEKRR:TYFSHLTPF
KQSAVREMAIVYKAAVLDKWKHYITGRKFVHIDORSKFLBOQVNMXYQWML:G
IGCDMIVYKPGVANKAADGLSR:QHPVYAILLA:TVPVYIQLODLYKEDEHEHIS
VIKQHDNLOGSHVC:VNGRIAYKOLLVFPKSSFIPLIOEYRDWEDISVDFTE
GLPTSGGVN:VLVVDRLSKAVYELGLKHLFKAJDVANKVSEVVK:YGFPEKSI:VDR
DRIELGAVKDKFSLGSLQKHSYSRPTDGTQ:EVNLRCLETYLRKFTSSHPRHKK
YLAWERFSNTAYSHAKTSFPKVVYGRDPPSLVRFELSGEDNDWJEVLKFLKZMFE
HIRANLQRAQEMKRNADMKRRDVFVAGDMVHLKLPYKQRTVVRPCYKLAAMFPG
PEFVIRGNVAYRINLPASSKIHNFVNSQLKVFGEQVQVQATNPPIKEVEDEFLP
EKIVDISFTKGNREFLVKWDORGVSENSWMPKFEVQVFPQFKLEDKLLFLARSDK
IHEAYVRQRTGPRQHLAVEGEDVAVTPPSLTIRNSAKDPGFQVLQFSSGHLPTSC
TA"
complement(47812..49281)
/gene="T10P12.4"
complement(join(47812..47877,46946..48223,48314..48405,
48706..48828,49010..49281))
/gene="T10P12.4"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39273.1"
/db_xref="GI:5080763"
/translacion="MAISLATAYISPOFTPESSNSASPSKTLSSVRLPSSQIRFGSVQ
SPSSIRFAPLTVSSPTLFLRDAIDFETKFNFLVLVTSSSFMYEANKLADGLAN
YAFSLPLGLHLESRPDVSLLADVAGVRAAKQIENSFDD:LONSDKPLVDVFA
TWQPCOLMVLINVESTLKDIAVYKIDTEKYPISLANKYQICALPIFILRQGLW
DRFEGALPANOLVER:ENSQVQK"
complement(50976..52602)
/gene="T10P12.5"
complement(join(50976..51522,51912..52404,52575..52602))
/gene="T10P12.5"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39277.1"
/db_xref="GI:5080767"
/translacion="MCLAQMTISOLPVYLGJELMTKAMTAHDYLLPLIKRKETS
WTGRFLSYGGRLOLTKSVMSITNWSAPFLPNCMKTERLCOSAFIWSQD:KTHN
AKIAWSKVLPMKEGGLGLRPLKLNVCCK:K:TWRLASOTSLSGQWVYLLIRNN
FWAKASQSGSWCHWFOATPKFAFTINLGNHRLSIGDRQKWNQALSTCVFQDQ
PLEIKDHLFFCHXANWIF:IAKENGVOYTSNMQDLASLIAG:SLTEPFCFCSML
FKPFTQSRESEITVGMKEQEQISNNRKPSTKQYTESPFSAREIKNMKTSSEVNS
PLEANLSITENFLFLKCC"
complement(60126..62065)
/gene="T10P12.6"
complement(join(60126..60522,60607..60800,60875..61272,
61788..62065))
/gene="T10P12.6"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39274.1"
Query Match 19.5%; Score 330.8; DB 8; Length 95190;
Best Local Similarity 78.3%; Pred. No. 3.7e-48;
Matches 503; Conservative 0; Mismatches 67; Indels 72; Gaps 6;
```

```
Qy 1 TGGGTTTTTGGGTAACATGACAAATATTATTATTATTTCATGAGTCTTTTATTCGATAG 60
|||||
Db 68585 TGGGTTTTTGGGTAACATGACAAATATTATTATTATTTCATGAGTCTTTTATTCGATAG 68526
```

```
Qy 61 CATGACAAATATTAAATATATATACAGTGTGTAATAACATGTTTGTCTTAAATACATGCAI 120
|||||
Db 68525 CATGACAAATATTAAATATATATACAGTGTGTAATAACATGTTTGTCTTAAATACATGCAI 68466

Qy 121 TTAA-AATCAGACATTTGTTTTTAAATCAAAATCTCTTATATCACAACGACATTGA 179
|||||
Db 68465 TTAGAATCATACATTTGTTTTTAAATCATATCTCT--TATCACAACGACATGA 68408

Qy 180 CGGAAAATCAGGTAAAGAGAAAATAAGAAATGAGAGATAGAGAGATTCTATGCGAA 239
|||||
Db 68407 TGGAGAATTCAGGCAAAAGAGAAAAGAGAAATGAGAGATAGAGAGATTCTATGCGAA 68348

Qy 240 AAGAAAGAGA-GAACAATGAGTGGTGAACAAATAAGAGATATGATATATATATGAG 298
|||||
Db 68347 AAGAAAGATAGGAACATGAGTGGTGAACAAATAAGAGATATGATATATATATGAG 68288

Qy 299 AGCTGGTCAAGATTAATTAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGGT 358
|||||
Db 68287 AGCTGGTCAAGATTAATTAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGGT 68229

Qy 359 ATCTGAA--GAAGATGAATTTGTTAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 415
|||||
Db 68228 AACTGGAAGATGAAGATGAGTTGTTCAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 68169

Qy 416 AAGTCTCGTAAAGAGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 473
|||||
Db 68168 AAGTCTC----- 68162

Qy 476 ATAGACTAATTAATGCGCAAAATTTCTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
|||||
Db 68161 -----ATTACCAAAATTTCTCTACAGACTAGTACTATTTGTTGTTCAAGGTTATT 68113

Qy 536 TTGCTATCTTTTGAAGTCAAAAGTTATTTCTTACATATATCTTCAAAATATAGAGAGAGAG 595
|||||
Db 68112 TTGTTATCTTTTCAAGTCAAAATGTTATTTGTAATATATCTATTTTCTTATATAT 68053

Qy 596 ACCAATTTTCCACACATGAGCTCTCTTTTATTCACAAAGTCA 637
|||||
Db 68052 ATATATATATATATATATAGAAITTTAATTTTAGAATTA 68011

RESULT 13
LOCUS 166494 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE : (bases 1 to 7218)
AUTHORS Dörner, F., Schefflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
SOURCE 1..7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 366 others
ORIGIN

Query Match 6.1%; Score 103.2; DB 6; Length 7218;
Best Local Similarity 13.2%; Pred. No. 1.6e-08;
Matches 63; Conservative 24; Mismatches 173; Indels 0; Gaps 0;

Qy 5 TTTTATTTGGATAACATGACAAATATTATTATTTCATGAGTCTTTTATTCGATAGCATG 64
|||||
Db 1593 TATATTTTACATACATATATCTTAATATGAAAGTGGTGAATCTGACTAGCTAGCATC 1534

Qy 65 ACAATATTAATATCATAGTCTTTAATACATGTTTGTCTTAAATATAGCATGCAITTTAA 124
|||||
Db 1533 GCCTCTAGACCATCTTATTCAGTTTCAAAAACCGCATGTAGGATCATCTATTAATTA 1474
```


* 34188 50517: contig of 15330 bp in length
 * 50518 55617: gap of 100 bp
 * 50618 65071: contig of 14454 bp in length
 * 65072 65171: gap of 100 bp
 * 65172 81054: contig of 15883 bp in length
 * 81055 81154: gap of 100 bp
 * 81155 95064: contig of 13930 bp in length
 * 95065 95184: gap of 100 bp
 * 95185 133197: contig of 38013 bp in length
 * 133198 133297: gap of 100 bp
 * 133298 171763: contig of 38466 bp in length.

FEATURES

Source
 Location/Qualifiers
 1..171763
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP24-358G9"
 /clone_lib="RPC1-24 Male Mouse BAC"

misc_feature
 1..1159
 /note="assembly_fragment"
 misc_feature
 1260..2312
 /note="assembly_fragment"
 misc_feature
 2413..3453
 /note="assembly_fragment"
 misc_feature
 3554..5119
 /note="assembly_fragment"
 misc_feature
 5220..7819
 /note="assembly_fragment"
 misc_feature
 7920..9949
 /note="assembly_fragment"
 misc_feature
 10050..14159
 /note="assembly_fragment"
 misc_feature
 14260..18550
 /note="assembly_fragment"
 misc_feature
 18651..25307
 /note="assembly_fragment"
 misc_feature
 25408..34087
 /note="assembly_fragment"
 misc_feature
 34188..50517
 /note="assembly_fragment"
 misc_feature
 50618..65071
 /note="assembly_fragment"
 misc_feature
 65172..81054
 /note="assembly_fragment"
 misc_feature
 81155..95084
 /note="assembly_fragment"
 misc_feature
 95185..133197
 /note="assembly_fragment"
 misc_feature
 133298..171763
 /note="assembly_fragment"

BASE COUNT 42403 a 34526 c 33542 g 53768 t 1524 others

ORIGIN

Query Match 538; Score 90.2; DB 2; Length 171763;
 Best Local Similarity 56.1%; Pred. No. 1,6e-06;
 Matches 170; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
 QY 194 AAAAGAGAAATATAAGATGAGAGATGAGAGATTTCATGCAAAAGAAAGAGAGAAC 253
 Db 130691 AGAGAGGGGGGAGGGAGGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 130692
 QY 254 ATGAGGTGAACAAATAAGAGATATGATGATATATTTATGAGAGGTGCTGTAAGATT 313
 Db 130631 AAGAAGAGAGAAAGAGAGAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGA 130632
 QY 314 TTTTAGAGAGGGAGAGAGAAATAGAAAAGAAATGATGCTGIGTAATCTGAGAGAGT 373
 Db 130571 GAGAAAGA 130572
 QY 374 AATTGTGTTAAAGTAGAT 433
 Db 130511 AAGAGAGAAAGA 130512
 QY 434 GAAAAGAGAAAGAT 493

Db 130451 AGAAGAGAAAGA 130452
 QY 494 AAA 496
 Db 130351 AAA 130389

RESULT 15

AC125184 280982 bp DNA linear HTG 23-JUN-2002
 Mus musculus chromosome UNK clone RP23-299L9, WORKING DRAFT
 LOCUS
 DEFINITION
 SEQUENCE, 4 unordered pieces.

AC125184 Gi:21490707
 AC125184.1 HTG: HTGS_PHASE1; HTGS_DRAFT.
 VERSION
 KEYWORDS
 SOURCE
 house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 McPherson,J.D. and Waterston,R.H.
 The sequence of Mus musculus clone
 Unpublished

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 2 (bases 1 to 280982)
 McPherson,J.D. and Waterston,R.H.
 Direct Submission
 Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 3 (bases 1 to 280982)
 McPherson,J.D. and Waterston,R.H.
 Direct Submission
 Submitted (23-JUN-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submission@watson.wustl.edu
 ----- Project Information -----
 Center project name: M_BA0299L09
 ----- Summary Statistics -----
 Sequencing vector: M13: 0%
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 277729 bases at least Q40
 Consensus quality: 279043 bases at least Q30
 Consensus quality: 279926 bases at least Q20
 Insert size: 197000; agarose-fp
 Insert size: 281848; sum-of-contigs
 Quality coverage: 0.00 in Q20 bases; agarose-fp
 Quality coverage: 9.94 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 8161: contig of 8161 bp in length
 * 8261: gap of unknown length
 * 51021: contig of 42760 bp in length
 * 6262
 * 51022 51121: gap of unknown length
 * 51122 84880: contig of 33759 bp in length
 * 84881 84980: gap of unknown length
 * 84981 280982: contig of 196002 bp in length.
 Location/Qualifiers

FEATURES

```
source
1. .280982
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-299L9"
1. .816:
misc_feature
/note="assembly_name:Contig10"
8262..51021
misc_feature
/oto="assembly_name:Contig11"
51122..84880
misc_feature
/note="assembly_name:Contig12"
84981..280982
misc_feature
/note="assembly_name:Contig13"
305 others
BASE COUNT 90913 a 51036 c 51472 g 87256 t
ORIGIN
Query Match 5.2%; Score 89.2; DB 2; Length 280982;
Best Local Similarity 54.6%; Pred. No. 2.2e-06;
Matches 178; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
QY 138 TTTTAAATCABAATCTAACTCTTATATACACACGACATTTGACGGNAATTCAGGGTAAAA 197
Db 162899 TTTTAACTGAAAAGAGCAACCTCATCAAAAAGAGAGAAAGCAAGAGAGAGAGAGAGA 162840
QY 198 AGAGAAAATAAGAGATGAGAGATAGAGAGATTTCTATGCAAAAAGAGAGAGACATCT 257
Db 162839 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 162780
QY 258 AGGTGAACAAATTAACAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTATTT 317
Db 162779 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 162720
QY 316 ACGAGAGGGAGAGAGAGAAATAGAAAAGAAAATGCACATGCTGAATCTGAAGAAGATCAAT 377
Db 162719 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 162660
QY 378 GTCTTAAAGATGAAGAGAGAGAGAGAACTCCATGGCTAAAGTCTGGTAAAGAGATGAAA 437
Db 162659 GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 162600
QY 438 AAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 493
Db 162599 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 162574
```

Search completed: December 4, 2002, 09:36:03
Job time : 6073 secs

GenCore version 5.1.3
 Copyright (c) 1993 - 2002 CompuGen, Ltd.
 OM nucleic - nucleic search, using sw model
 Run on: December 4, 2002, 06:40:33 : Search time 271 Seconds
 (without alignments)
 14125.927 Million cell updates/sec

Title: US-09-733-685-3
 Perfect score: 1700
 Sequence: 1 tgggtttttatggaaca.....tttcgacttacgaacata: 1700

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 2145239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370476

Minimum DB seq length: 3
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 04
 Maximum Match: 100%
 Listing first 45 summaries

Database : N_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	Description
C 1	76.8	4.5	7657 22	AAS45477 Chemically pretreated
C 2	76.8	4.5	7657 24	ABJ34022 Human immune syste
C 3	76.8	4.5	14006 24	ABJ33958 Human immune syste
C 4	76	4.5	19124 18	AA72882 Plasmodium var-7 g
C 5	76	4.5	19124 21	AA72882 Plasmodium var-7 p
C 6	74.8	4.4	9539 22	AAS45347 Chemically pretrea
C 7	74.8	4.4	9539 24	ABK28180 DNA transcription
C 8	74	4.4	6644 20	AAJ33181 Base sequence of t
C 9	74	4.4	7372 20	AAJ33182 Base sequence of t

10	74	4.4	7797 20	AAJ33180 Cowpox virus bsr f
11	74	4.4	7996 20	AAJ33184 Base sequence of t
C 12	73.4	4.3	113515 24	ABJ34174 Human immune syste
C 13	72.6	4.3	16033 24	ABJ33404 Human immune syste
C 14	72	4.2	16167 24	ABJ70254 Chemically treated
C 15	72	4.2	16167 24	ABJ33083 Human immune syste
C 16	72	4.2	16167 24	ABJ34529 Human metastasis a
C 17	71.6	4.2	17934 24	ABJ33719 Human immune syste
C 18	71.2	4.2	10620 20	AAJ22996 Human H-1ra BAC c
C 19	71.2	4.2	14690 20	AAJ22303 Human H-1ra BAC c
C 20	70.6	4.2	34769 22	AAS46775 Tumour suppressor
C 21	70	4.1	5750 22	AAS46708 Tumour suppressor
C 22	70	4.1	5750 24	ABJ34008 Human immune syste
C 23	69.6	4.1	556 24	ABQ36996 Oligonucleotide fo
C 24	69.6	4.1	556 24	ABQ36997 Oligonucleotide fo
C 25	69.4	4.0	723 24	ABQ25178 Oligonucleotide fo
C 26	69.4	4.0	723 24	ABQ25179 Oligonucleotide fo
C 27	68	4.0	292 24	ABJ86605 Human ovarian canc
C 28	68	4.0	6145 24	ABJ32972 Human immune syste
C 29	68	4.0	16633 24	ABJ79984 Human chemically m
C 30	67.8	4.0	8079 24	ABJ92313 Chemically treated
C 31	67.4	4.0	8404 22	AAJ46500 Tumour suppressor
C 32	67.4	4.0	8404 24	ABJ33595 Human immune syste
C 33	67.2	4.0	556 23	ABJ40063 Human prostate exp
C 34	67.2	4.0	556 23	ABJ40163 Human prostate exp
C 35	67.2	4.0	556 23	ABJ42105 Human prostate exp
C 36	67.2	4.0	556 23	ABJ43601 Human prostate exp
C 37	67	3.9	6668 24	ABJ33697 Human immune syste
C 38	66.8	3.9	174424 24	ABJ68122 Ovary cancer relat
C 39	66.6	3.9	6485 22	AAS46559 Tumour suppressor
C 40	66.6	3.9	6485 24	ABJ33808 Human immune syste
C 41	66.4	3.9	299 24	ABJ86618 Human ovarian canc
C 42	66.4	3.9	563 23	ABJ50724 Human prostate exp
C 43	66.4	3.9	6338 24	ABJ33126 Human immune syste
C 44	66.4	3.9	16200 24	AAD36438 Mouse L66 genomic
C 45	66.2	3.9	307 24	ABJ86438 Human ovarian canc

ALIGNMENTS

RESULT :

10 AAS45477/C
 AA45477 standard; DNA; 7657 BP.

XX AAS45477:

XX 18-DEC-2001 (first entry)

XX Chemically pretreated complementary DNA associated with cell cycle #91.

XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;

XX human immunodeficiency virus; neurodegenerative disorder; solid tumour;

XX graft-versus-host disease; glomerular disease; Lewy body disease; cancer;

XX arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;

XX immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;

XX PCR primer.

XX Homo sapiens.

XX WC2000168911-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02945.

XX 15-MAR-2000; 2000DE-1013847.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPiGENOMICS AG.

```
XX
PI Olek A. Piepenbrock C, Berlin K;
XX
DR WPI: 2001-602751/68.
XX
PT Designing primers and probes for analysing diseases associated with
XX cytosine methylation state e.g. arthritis, cancer, aging,
XX arthritisclerosis comprising fragments of chemically modified genes
XX associated with cell cycle.
XX
XX Claim 1: SEQ ID NO 182; 28pp; English.
XX
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
XX molecules associated with the cell cycle and specific PCR primers of the
XX invention. The sequences are useful for detecting the methylation state
XX of all CpG dinucleotides in a sequence and therefore for analysing
XX associated diseases. By analysing cytosine methylations in the pretreated
XX DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
XX of existing diseases or the predisposition to specific diseases can be
XX ascertained. The parameters may be compared to another set of genetic
XX and/or epigenetic parameters, the differences serving as basis for
XX diagnosis and/or prognosis events which are disadvantageous to patients.
XX The sequences of the invention are useful for the diagnosis and therapy
XX of HIV infection, neurodegenerative disorders, graft-versus-host disease,
XX aging, g-omerular disease, Lewy body disease, arthritis,
XX arthritisclerosis, solid tumours and cancers.
XX
XX Sequence 7657 BP; 2028 A; 82 C; 1628 G; 3919 T; 0 other;
XX
XX Query Match 4.5%; Score 76.8; DB 22; Length 7657;
XX Best Local Similarity 48.6%; Pred. No. 6e-06;
XX Matches 210; Conservative 0; Mismatches 222; Indels 0; Gaps 0;
XX
QY 65 ACAATAATTATATACATGTCATACATGTTTCTTTTAAATACATGATTTTAA 124
DB 6067 AAAAAATTAATATACACATCAAAATTAATTAACCTATAAAATAAAAAAATCAATT 6068
QY 125 AATCAGACATTTGTTTAAATCAAAATCTATCTTATATACACGACATTCACGGAA 184
DB 6007 AACTCCAAAAAACAACAAATATCCCGACATTAATATATATACCCCAAAAAAACGTA 5948
QY 185 AATTCAGGTAAAAGAGAAATAAAGATGAGATGAGAGATTTCTATUGAAAAAGAA 244
DB 5947 AACGAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 5888
QY 245 AGAGAGACATGTAGTGAACAAAAATAAGAGATGATGATATATTTTATGAGAGTGG 304
DB 5887 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAA 5828
QY 305 TGAAGATTATTTTAGGAGAGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTG 364
DB 5827 AAAAAATTAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 5768
QY 365 AAGAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAACTCCATGGCTTAAGTCTCGT 424
DB 5767 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAA 5708
QY 425 AACAAGATGAARAGAAACAAAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484
DB 5707 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAA 5648
QY 485 CTATGGCCAAA 496
DB 5647 ACTTACTAATA 5636
```

RESULT 2

ABL34022/c

ID ABL34022 standard; DNA; 7657 BP.

XX

AC ABL34022;

XX

DT 26-MAR-2002 (first entry)

```
XX
DE Human immune system associated gene SEQ ID NO: 1995.
XX
XX Human: immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytostatic; nootropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritis; antidiabetic; antipsoriatic;
XX antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene: ds.
XX
XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPig-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI: 2302-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation.
XX
XX Claim 1: SEQ ID NO 1995; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX
XX Sequence 7657 BP; 2028 A; 82 C; 1628 G; 3919 T; 0 other;
XX
XX Query Match 4.5%; Score 76.8; DB 24; Length 7657;
XX Best Local Similarity 48.6%; Pred. No. 6e-06;
XX Matches 210; Conservative 0; Mismatches 222; Indels 0; Gaps 0;
XX
QY 65 ACAATAATAATATATACATGTTTAAATCAAAATCTTGTCTTAAATACATGATTTTAA 124
DB 6067 AAAAAATTAATATACACATCAAAATTAATTAACCTATAAAATAAAAAATAAAAAA 6068
QY 125 AATCAGACATTTGTTTAAATCAAAATCTTATCTTATATACACGACATTCACGGAA 184
DB 6007 AACTCCAAAAAACAACAAATATCCCGACATTAATATATATACCCCAAAAAAACGTA 5948
QY 185 AATTCAGGTAAAAGAGAAATAAAGATGAGATGAGAGATTTCTATGAAAAAGAA 244
DB 5947 AACGAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAA 5888
QY 245 AGAGAGACATGTAGTGAACAAAAATAAGAGATGATGATATATTTTATGAGAGTGG 304
DB 5887 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAA 5828
QY 305 TGAAGATTATTTTAGGAGAGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTG 364
DB 5827 AAAAAATTAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 5768
QY 365 AAGAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAACTCCATGGCTTAAGTCTCGT 424
DB 5767 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAA 5708
```


PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and sialic acid
 PT binding proteins

XX Claim 4: Page 56-61: 96pp; English.

XX This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and sialic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions are for the treatment and prevention of
 CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 CC family of genes having homology with conserved regions of DABP and SABP.
 CC The compositions are used for the treatment and prevention of malaria.
 CC They are also used in the preparation of vaccines for inducing a
 CC protective immune response in a mammal to Plasmodium merozoites
 CC (especially Plasmodium falciparum or Plasmodium vivax).

XX Sequence 19124 BP; 7624 A; 2190 C; 2790 G; 6320 T; 0 other;

Query Match 4.5%; Score 76; DB 18; Length 19124;
 Best Local Similarity 50.0%; Pred. No. 1e-05;
 Matches 218; Conservative 0; Mismatches 215; Indels 3; Gaps 1;
 QY 52 ATTGGATGACGACAAATATTAAATATATCAGTGTAAATACATGTTTGTCTTAAAT 111
 DB 15440 ATTATTAANAAGAAAAGAAAATGAAATATTAANAANAATATTAAANAATAAANA 15499
 QY 112 ACATGCAATTTAAATCAGACATTTGTTTAAATCAAAATCTATATATCACAAC 171
 DB 15500 AAAAAAAGAAAAGGAGAAAATTTTAAATAATATATAAATAATATAAATAATA 15559
 QY 172 GACATTCAGCGAAATTCAGGTAAAAGAGAAATAGAAATGAGAGATGAGAGATTTC 231
 DB 15560 AATTTGATAGATAAANAATGAAAAGAGATTATCAANAANAANAATTAANAANA 15616
 QY 232 TATCGAAAGAAAGAGAGACATGTAGGTGAACAAAATAAAGAGATATGATGATATAT 291
 DB 15617 TATATAAANAANAATGATTATTAANAANAATTAANAANAATTAANAANAATA 15676
 QY 292 TATGAGAGGTGCTGAAGATGATTTAGGAGGGGAGAGAGAAATAGAAAAGAAATGA 351
 DB 15677 AAAAAAAGAAAAGGAGAAAATTTTAAATAATATATAAANAATTAANAANAATA 15736
 QY 352 CATGGTGAATCTGAAGATGAATTCGTTAAGATGAGAGAGAGAGAACTCCCATG 411
 DB 15737 AAATATATCATTAANAANAANAATTAANAANAATTTAAANAANAATTAANAANA 15796
 QY 412 GCTAAAGTCTGTAAGAGAGATGAAAAGAGAAACAAAGAGAGAGAGAAAGGC 471
 DB 15797 AAATAAANAANAATTAATTAANAANAANAATTAANAANAATTAANAANAATA 15856
 QY 472 TAAATAGACTACTA 487
 DB 15857 AAAAAAATAAATAA 15872

RESULT 5

AAZ98287 standard; DNA; 19124 BP.

XX AAZ98287;

XX 13-JUN-2000 (first entry)

XX

DE Plasmodium var-7 polypeptide encoding DNA.

XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 XW DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
 XW protozoan; var-7; ds.

XX Plasmodium sp.

XX CS5999327-A.

XX 30-NOV-1999.

XX 07-JUN-1995; 95US-0487826.

XX 10-SEP-1993; 93US-0119677.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Sim KL, Chitnis C, Peterson DS, Su X, Wellem TE, Miller LH;

XX WPI; 2000-194198/17.

XX P-PSDB; AAY77904.

XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria.

XX Disclosure; Columns 91-108; 93pp; English.

XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunological studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria. The present sequence represents the var-7 polypeptide encoding
 CC DNA.

XX Sequence 19124 BP; 7624 A; 2190 C; 2790 G; 6320 T; 0 other;

Query Match 4.5%; Score 76; DB 21; Length 19124;
 Best Local Similarity 50.0%; Pred. No. 1e-05;
 Matches 218; Conservative 0; Mismatches 215; Indels 3; Gaps 1;

QY 52 ATTGGATGACGACAAATATTAAATATATCAGTGTAAATACATGTTTGTCTTAAAT 111
 DB 15440 ATTATTAANAAGAAAAGAAAATGAAATATTAANAANAATTTAAANAATAAANA 15499
 QY 112 ACATGCAATTTAAATCAGACATTTGTTTAAATCAAAATCTATATATCACAAC 171
 DB 15500 AAAAAAAGAAAAGGAGAAAATTTTAAATAATATATAAATAATATAAATAATA 15559
 QY 172 GACATTCAGCGAAATTCAGGTAAAAGAGAAATAGAAATGAGAGATGAGAGATTTC 231
 DB 15560 AATTTGATAGATAAANAATGAAAAGAGATTATCAANAANAANAATTAANAANA 15616
 QY 232 TATCGAAAGAAAGAGAGACATGTAGGTGAACAAAATAAAGAGATATGATGATATAT 291
 DB 15617 TATATAAANAANAATGATTATTAANAANAATTAANAANAATTAANAANAATA 15676
 QY 292 TATGAGAGGTGCTGAAGATGATTTAGGAGGGGAGAGAGAAATAGAAAAGAAATGA 351
 DB 15677 AAAAAAAGAAAAGGAGAAAATTTTAAATAATATATAAANAATTAANAANAATA 15736
 QY 352 CATGGTGAATCTGAAGATGAATTCGTTAAGATCAAGAGAGAGAGAACTCCCATG 411
 DB 15737 AAATATATCATTAANAANAANAATTAANAANAATTTAAANAANAATTAACA 15796
 QY 412 GCTAAAGTCTGTAAGAGAGATGAAAAGAGAAACAAAGAGAGAGAGAAAGGC 471

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: December 4, 2002, 06:41:38 : Search time 71 seconds
(without alignments)
9336.418 Million cell updates/sec

File: US-09-733-685-3
Perfect score: 1700
Sequence: 1 tgggtttttattgataaca.....tttagcacttaagacatat 1700

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 352425 seqs, 194366365 residues
Total number of hits satisfying chosen parameters: 750850

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCT_US_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	73.8	4.3	529	10	US-09-983-965-2109 Sequence 2109, Ap
2	72	4.2	446	10	US-09-960-352-3400 Sequence 3400, Ap
3	71.2	4.2	152331	9	US-10-095-407-16 Sequence 16, Appl
4	71.2	4.2	176373	9	US-10-095-407-17 Sequence 17, Appl
5	68	4.0	292	10	US-09-867-701-9583 Sequence 9583, Ap
6	68	4.0	516	10	US-09-960-352-5785 Sequence 5785, Ap
7	67.8	4.0	180216	10	US-09-835-232-6 Sequence 6, Appl
8	66.8	3.9	174424	10	US-09-967-768A-314 Sequence 314, Appl
9	66.4	3.9	299	10	US-09-867-701-9596 Sequence 9596, Ap
10	66.2	3.9	307	10	US-09-867-701-9416 Sequence 9416, Ap
11	65.6	3.9	42999	9	US-09-799-462A-17 Sequence 17, Appl
12	65.6	3.9	42999	9	US-10-125-767-17 Sequence 17, Appl
13	65.2	3.8	451	10	US-09-960-352-10262 Sequence 10262, A
14	63.8	3.8	174424	10	US-09-967-768A-314 Sequence 314, Appl
15	63.4	3.7	428	10	US-09-960-352-573 Sequence 573, Appl
16	63	3.7	299	10	US-09-867-701-9477 Sequence 9477, Ap
17	63	3.7	341	10	US-09-960-352-12302 Sequence 12302, A
18	62.6	3.7	152331	9	US-10-C95-407-16 Sequence 16, Appl
19	62.4	3.7	1721	9	US-09-938-842A-3578 Sequence 3578, Ap

20	62.4	3.7	4157	10	US-09-764-878-282 Sequence 282, App
21	62.4	3.7	4157	10	US-09-764-860-1145 Sequence 1145, Ap
22	62.4	3.7	4157	10	US-09-764-846-345 Sequence 345, App
23	62	3.6	425	10	US-09-960-352-4010 Sequence 4010, Ap
24	62	3.6	8317	10	US-09-764-869-1279 Sequence 1279, Ap
25	61.8	3.6	344	10	US-09-960-352-1036 Sequence 1036, Ap
26	61.8	3.6	431	10	US-09-960-352-5558 Sequence 5558, Ap
27	61.4	3.6	26555	9	US-09-860-670-161 Sequence 161, Appl
28	61.2	3.6	90341	10	US-09-759-359A-3 Sequence 3, Appl
29	60.4	3.6	463	10	US-09-960-352-7186 Sequence 7186, Ap
30	60	3.5	419	10	US-09-960-352-11234 Sequence 11234, A
31	59.2	3.5	393	10	US-09-960-352-4582 Sequence 4582, Ap
32	58.4	3.4	408	10	US-09-960-352-1221 Sequence 1221, Ap
33	57.8	3.4	172637	10	US-09-805-458A-3 Sequence 3, Appl
34	56.8	3.3	659158	9	US-09-771-208-20 Sequence 20, Appl
35	56.2	3.3	298	10	US-09-867-701-8978 Sequence 8978, Ap
36	56	3.3	42999	10	US-09-740-029-3 Sequence 3, Appl
37	55.8	3.3	545	10	US-09-878-574-4299 Sequence 4299, Ap
38	55.8	3.3	1267	12	US-10-001-843-45 Sequence 45, Appl
39	55.6	3.3	277	10	US-09-867-701-8971 Sequence 8971, Ap
40	54.8	3.2	368	10	US-09-834-975-47 Sequence 47, Appl
41	54.8	3.2	439	10	US-09-960-352-1009 Sequence 1009, Ap
42	54.6	3.2	349	10	US-09-867-701-9513 Sequence 9513, Ap
43	54.6	3.2	423	10	US-09-864-761-18355 Sequence 18355, A
44	54.6	3.2	465237	10	US-09-933-267A-1 Sequence 1, Appl
45	54.6	3.2	640581	10	US-09-790-988-1 Sequence 1, Appl

ALIGNMENTS

RESULT :
US-09-983-965-2109/C
Sequence 2109, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Mengbing
APPLICANT: Svatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 05/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 63/113,678
PRIOR FILING DATE: 1996-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 2109
LENGTH: 529
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: unsure
LOCATION: (335)
OTHER INFORMATION:
OTHER INFORMATION: Clone ID: 34-LiB3057-015-Q1-K1-A6
US-09-983-965-2109

Query Match	4.3%	Score	73.8	DB	10	Length	529
Best Local Similarity	49.2%	Pred. No.	6.7e+06				
Matches	192	Conservative	0	Mismatches	198	Indels	0
Gaps	0						
Cy	107	AAATACATGCATTTTAAATCAGACATTTGTTTTAAATCAAAATCTAATCTCTATATC	166				
DB	474	AA	415				
Cy	167	ACACGACATTGACCGAAATTCAGGTAAAGAGAAATTAAGAAATCAGAGATAGAGAG	226				
DB	414	AA	355				

```
QY 227 ATTCTATGGAAGAAAGAGGACATGTAGTGGACCAAAATTAAGAGCATTAATGAT 286
DB 354 AAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAC 293
QY 287 ATATTATTACAGAGTGTTGAGAGATTATTAGGAGAGGAGAGAGAAATACAAAAGAA 346
DB 294 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 235
QY 347 ATGACATGGTGAATCTGAAGAGATGAATTTGTTAAAGATGGAAGAGAGAAAGAACT 406
DB 234 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 175
QY 407 CCATGCTTAAGTCTGTAAGAGAGTGAAGAGCAACAAAGCAAGAGGAGAGAAAGACA 466
DB 174 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 115
QY 467 AAGGCTAAATAGACTAATTTGCCAAA 496
DB 114 AAAAAAANAANAACCAACATAAAAAANA 85

RESULT 2
US-09-960-352-3400/c
; Sequence 3400, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Rao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.066/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; PRIOR FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3400
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 15-L1B305b-C52-Q1-K1-D11
US-09-960-352-3400

Query Match 4.2%; Score 72; DB 10; Length 446;
Best Local Similarity 48.5%; Pred. No. 1.4e-05;
Matches 198; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 108 AATACATGCAATTTTAAATCGACATTTTGTTTTAAATCAATCTAATCTCTTTATTA 167
DB 409 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 150
QY 158 CAACGACATTTGCGGAAATTCAGCTAAAGAGAGAAATTAACATGAGNGAAGAGAA 227
DB 349 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 290
QY 228 TTCTATGGAAGAAAGAGAGACATGTAGTGAACAAATAAAGACATATGATGATA 287
DB 289 AATAAATATAAANAATAAANAATAAANAATAAANAATAAANAANAANAANAANAANA 250
QY 288 TATTATTAGAGGTGCTGAAGATTATTTTAGAGAGGAGAGAGAAATAGAAAAGAAA 347
DB 229 AATATAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 170
QY 348 ATGACATGGTGAATCTGAAGACATGAATTTGTTTAAAGATGAAGAGAGAAAGCACTC 407
DB 169 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 110
QY 408 CATGGCTAAGTCTGCTAAAGAGAGTAAAAAGAAACAAAGAGGAGAGAAAGAGAA 467
DB 109 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 50
QY 468 AGGCTAAATAGACTAATTCGCCAAAATTTCTGTAGCGCAATA 515
DB 114 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 85
```

```
DE 45 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGGGCGGCTCGCATCTA 2
RESULT 3
US-10-095-407-16/C
; Sequence 16, Application US/10095407
; Patent No. US20020164330A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/10/095,407
; PRIOR FILING DATE: 2002-03-11
; PRIOR APPLICATION NUMBER: US 60/091,650
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/054,646
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152311
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(152311)
; OTHER INFORMATION: n = A,T,C or G
US-10-095-407-16

Query Match 4.2%; Score 71.2; DB 9; Length 152331;
Best Local Similarity 54.6%; Pred. No. 9.6e-05;
Matches 142; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 253 AAATAAAGAAATGACAGATAGAGATTTCTATGGAAGAAAGAGAGAGAACATGTAGG 262
DB 128751 AGACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 128652
QY 253 AACAATAAAGAGATATGATGATATTTTATGAGAGGTGGTGAAGTATTTTAGGAG 322
DB 128651 AGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128632
QY 323 AGGAGAGAGACAAATAGAAAAAAGAAATGACATGTGTAATCTGAAGAGAGATCAAT 382
DB 128631 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128572
QY 383 AAATATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 442
DB 128571 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128512
QY 443 ACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 462
DB 128511 ACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128492

RESULT 4
US-10-095-407-17/C
; Sequence 17, Application US/10095407
; Patent No. US20020164330A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/10/095,407
; PRIOR FILING DATE: 2002-03-11
; PRIOR APPLICATION NUMBER: US 60/091,650
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/054,646
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
```


[illegible]

QY 376 TTGTTTAAGATGAGAGAGAGAGAACTCCATGGCTAAAGTCTCGTAAGAGAGATCA 435
DB 190 AAACGAGCGAGGAGGAGGAGCGGAGAGCGGAGAGAAAGAAAGAAAGAAAGAA 249
QY 436 AAAAGAAACAAAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471
DB 250 GAAAGAAAGATC 285

RESULT 10

US-09-867-701-9416
; Sequence 9416, Application US/09867701
; Patent No. US2002013237A1
; GENERAL INFORMATION:
; APPLICANT: Agilate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701;
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9416
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9416

Query Match 3.9%; Score 66.2; DR 10; Length 307;
Best Local Similarity 52.8%; Pred. No. 0.00017;
Matches 143; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
QY 198 AGAGAAATTAAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATG 257
DB 23 ACAGAGCGGAG 82
QY 258 AGGTGACAAATAAAGAGATATGATGATATATTTATGAGAGAGTGTGAGATTTATTT 317
DB 83 GAGAGCGGAG 142
QY 318 AGGAGGGGAT 377
DB 143 GAGCAG 202
QY 378 GTGTTTAAAGATGAAG 437
DB 203 ACAGCGGAGGAG 262
QY 438 AAGAAACAAAG 468
DB 263 AAGAAACAAAG 293

RESULT 11

US-09-799-462A-17/C
; Sequence 17, Application US/09799462A
; Patent No. US20020160970A1
; GENERAL INFORMATION:
; APPLICANT: Radiacsky, Gylia
; APPLICANT: Szalay, Aladar
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McCulliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/799,462A
FILING DATE: 10-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/835,682
FILING DATE: 10-APR-1997
APPLICATION NUMBER: 08/695,191
FILING DATE: 07-AUG-1996
APPLICATION NUMBER: 08/682,080
FILING DATE: 15-JUL-1996
APPLICATION NUMBER: 08/629,822
FILING DATE: 10-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24601-402G
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858-450-8403
TELEFAX: 858-587-5350
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 42999 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: <Unknown>
ORIGINAL SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-799-462A-17

Query Match 3.9%; Score 65.6; DR 9; Length 42999;
Best Local Similarity 52.5%; Pred. No. 0.00085;
Matches 166; Conservative 0; Mismatches 149; Indels 1; Gaps 1;
QY 166 CACACGACATTCAGCGGAAATTCAGGTAAAAAGAGAGAAATAAGAGATGAGAGATAGAGA 225
DB 27215 CAGACAGACAGCGGGGAGAGAAAGACAAACAGAGAGAGAGAGAGAGAGAGAGAGAGA 27156
QY 226 CATTTCTATGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATATGTA 285
DB 27155 GACACAGACAG 27096
QY 286 TAATTTTATGAGAGTGTGTAAGATTTATTTTAGGAGGAGGAGAGAGAGAGAGAGAGAG 345
DB 27095 GAGAGAGATGAG 27036
QY 346 AATGACATGGTGAATCTGAAGAGAGATGAATTTGTTTAAAGATGAAGAGAGAGAGAGAG 405
DB 27035 AACAGCAG 26976
QY 406 TCCATGGCTAAAGTCTCGTAAAGA-AGATGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 464
DB 26975 TGGAGGGAG 26916
QY 465 GAAAGGCTAAATAGA 480
DB 26915 GTGAGTGAAG 26900

RESULT 12

US-10-125-767-17/c
; Sequence 17, Application US/10125767
; Patent No. US20020160410A1

[illegible]

PRIOR FILING DATE: 2000-09-28

NUMBER OF SEQ ID NOS: 325

SOFTWARE: PatentIn version 3.0

SEQ ID NO 314

LENGTH: 174424

TYPE: DNA

ORGANISM: Homo sapiens

US-09-967-768A-314

Query Match 3.8% Score 63.8; DB 10; Length 174424;
Best Local Similarity 53.2%; Pred. No. 0.0028;
Matches 157; Conservative 0; Mismatches 137; Indels 1; Gaps 1;
QY 190 AGCTAAAGAGCAAAATAAGCAATGAGAGATAGAGAGATTCTTATGGAAAGAGAGAGA 249
DB 13725 AAGAAGAAAGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 13666
QY 250 GAACATGTAGTGAACAAATAAGAGATATGATATATTTATGAGAGGTGGTGAAG 309
DB 13665 GCGAGGGAGGGAGGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 13606
QY 310 ATTATTTAGGAGAGGA 369
DB 13605 GAGGAGGGGAGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 13547
QY 370 GATGATATGTTTAAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
DB 13546 AAGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 13487
QY 430 AGATGAAAGAA 484
DB 13486 AAGAA 13432

RESULT 15

US-09-960-352-573/C
Sequence 573, Application: US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Marion, Wesley C.
APPLICANT: Tao, Mengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Naganathan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16S11.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 573
LENGTH: 428
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 03-L1B3058-031-Q1-K1-A11
US-09-960-352-573

Query Match 3.7% Score 63.4; DB 10; Length 428;
Best Local Similarity 52.8%; Pred. No. 0.0067;
Matches 159; Conservative 0; Mismatches 141; Indels 1; Gaps 1;
QY 196 AAGAGAGAAATAAGCAATGAGAGATAGAGAGATTCTTATGGAAAAAGAGAGAGAGAGAT 255
DB 428 AAAAAATAAATAAATAATAATAA-AAGAAATAAATAAATAAATAAATAAATAAATAA 170
QY 256 GTAGGTGACAGCAAAATAAGAGATATGATATATTTATGAGAGGTGGTGAAGATATT 315
DB 369 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 310
QY 316 TTAGGAGAGGAA 375
DB 309 AAGAAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 250
QY 376 TTGTGTTAAAGATGAAGATGA 435

DB 249 AAAAAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 190
QY 436 AAAAAAACAAGAA 495
DB 189 AAAAAAGAA 130
QY 496 A 496
DB 129 A 129

Search completed: December 4, 2002, 07:17:49
Job time : 1414 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Ran on: December 4, 2002, 07:05:43 : Search time 2349 Seconds
(without alignments)
13436.936 Million cell updates/sec

Title: US-09-733-685-3
Perfect score: 1700
Sequence: 1 tgggttttttttgcgaacaa.....tttcagacttgcgaacatat 1700

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs. 8097743376 residues

Total number of hits satisfying chosen parameters: 3208132

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 5%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_estbur:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estom:*
 - 6: em_estpl:*
 - 7: em_estio:*
 - 8: em_hic:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: gb_oss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vit:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_fod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	230	13.5	230	17	BH866477
2	203.8	12.0	531	17	BH866479
3	201.2	11.8	206	17	BH866474
4	185	10.9	237	17	BH814587
5	175.2	10.3	196	17	BH866500
6	173.6	10.2	267	17	BH854168

C	7	128.4	7.6	387	17	BH749275
C	8	101.6	6.0	1101	17	CNS0021J
C	9	96	5.6	660	17	BH183498
C	10	96	5.6	660	17	CNS070N
C	11	91	5.4	982	17	AQ325799
C	12	89	5.2	529	9	AL514657
C	13	88	5.2	769	17	CNS01V5E
C	14	86.6	5.1	1101	17	CNS008VL
C	15	86.4	5.1	764	17	AZ193158
C	16	85.8	5.0	842	17	CNS010QN
C	17	85.6	5.0	937	17	CNS005TE
C	18	85	5.0	886	17	CNS075DF
C	19	85	5.0	934	17	AZ184244
C	20	85	5.0	1201	17	CNS0167M
C	21	84.8	5.0	1101	17	CNS001T2
C	22	84	4.9	647	17	BH038922
C	23	84	4.9	839	14	BQ151187
C	24	83.8	4.9	922	17	AZ548363
C	25	83.4	4.9	636	17	AQ329262
C	26	83.4	4.9	846	17	AZ549808
C	27	82.6	4.9	89	17	BH866492
C	28	82.6	4.9	718	17	AG160666
C	29	82.4	4.8	434	17	CNS06Q23
C	30	82.4	4.8	952	17	AQ897460
C	31	82.2	4.8	580	17	CNS008HN
C	32	82.2	4.8	1101	17	CNS0039G
C	33	82	4.8	888	17	AZ549422
C	34	81.9	4.8	625	17	CNS036A2
C	35	81.8	4.8	915	17	AG044049
C	36	81.4	4.8	891	17	CNS009J9
C	37	81.2	4.8	1031	17	CNS0064G
C	38	81.2	4.8	1034	17	CNS06PLB
C	39	81.2	4.8	1101	17	CNS012TP
C	40	81	4.8	89	17	BH866501
C	41	80.8	4.8	1042	17	CNS0148K
C	42	80.6	4.7	797	10	BE034282
C	43	80.6	4.7	1101	17	CNS0161N
C	44	80.4	4.7	614	17	CNS0152H
C	45	79.8	4.7	987	17	CNS00418

ALIGNMENTS

RESULT 1
BH866477
LOCUS
DEFINITION SALK_101380 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_101380, DNA sequence.
ACCESSION BH866477
VERSION BH866477.1
KEYWORDS GSS.
SOURCE
ORGANISM Arabidopsis thaliana
thale cress.

REFERENCE 1 (bases 1 to 230)
Alonso,J.M., Leisse,I.J., Barajas,P., Chen,H., Cheek,R., Gadrinab

C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel.: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

BH866477 230 bp DNA linear GSS 05-AUG-2002
SALK_101380 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_101380, DNA sequence.

ACCESSION BH866477
VERSION BH866477.1
KEYWORDS GSS.
SOURCE
ORGANISM Arabidopsis thaliana
thale cress.
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae: Arabidopsis.
1 (bases 1 to 230)
Alonso,J.M., Leisse,I.J., Barajas,P., Chen,H., Cheek,R., Gadrinab
C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel.: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

High quality sequence stop: 660.

```

FEATURES
  source
    Location/Qualifiers
      1..660
        /organism="Schistosoma mansoni"
        /strain="Puerto-Rican"
        /db_xref="taxon:6183"
        /clone="023CP04"
        /sex="mixed"
        /dev_stage="cercariae"
        /lab_host="Blomphalaria glabrata"
        /note="Vector: pBeloBAC 11; Site: 1: Hind III; Partially
        Hind III digested and size-selected S. mansoni cercarial
        DNA was ligated into Hind III digested pBeloBAC 11 vector
        and used to transform E. coli DH10B. The complete library
        contains 23808 clones from 4 independent
        sizing-ligation-transformations. Average insert size
        ranges from 70-127 kb and genome coverage is 7.9-fold."

BASE COUNT      139 a   16 c   9 g   446 t   50 others
ORIGIN

Query Match      5.6%   Score 96;   DB 17;   Length 660;
Best Local Similarity 46.5%;   Pred. No. 1.3e-06;
Matches 225;   Conservative 32;   Mismatches 225;   Indels 2;   Gaps 1;

QY  15  ATAACATGACAAATATTATTTCATGAGTTTATTTCGATAGCATGCACAAATATTA 74
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  615  AAAAAAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 556

QY  75  ATATATCATGTTTAAATACATGTTTTCCTTATATCACAACGACATTCGCGAAATTCAGG 134
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  615  AAAAAAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 556

QY  135  TGGTTTAAATCAATCTAA--TCTCTTATATCACAACGACATTCGCGAAATTCAGG 192
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  495  AATWAAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 436

QY  193  TAAAGAGAGAAATTAAGAGATGAGAGATGAGAGATTTCTATGGAAAGAGAGAGAGAA 252
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  435  ATATWTATATATTAATAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 376

QY  253  CATGTAGTGCACAAATAAGAGATATGATATATTTATCAGAGGTGGTGAAGATT 312
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  375  GAAAAAATAATATAAAAAATGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 316

QY  313  ATTTAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 372
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  315  KAAAGGTATTCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 256

QY  373  GAATTGTGTTAAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  255  AAAAAAATAATATAAAAAATGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 196

QY  433  TGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  195  AAAAAAATAATATAAAAAATGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 136

QY  493  AAAAA 496
    |||
DB  135  AAAAA 132

RESULT 10
CNS070N1/C      660 bp   DNA   linear   GSS 10-OCT-2001
LOCUS
DEFINITION
  T3 end of clone 023CP04 of library SmBAC1 from strain Puerto-Rican
  of Schistosoma mansoni, genomic survey sequence.
ACCESSION
  AL620449
VERSION
  AL620449.1   GI:16034593
KEYWORDS
  GSS
SOURCE
  Schistosoma mansoni.
  Schistosoma mansoni
  Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

```

```

Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
1 (bases 1 to 660)
Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
Williams,D.L., Johnston,D., LoVerde,P.I. and Le Paslier,D.
Construction and characterization of a Schistosoma mansoni
bacterial artificial chromosome library
Genomics 65 (2), 87-94 (2000)
20247247
MEDLINE
PUBMED
10783255
REFERENCE
2 (bases 1 to 660)
Genoscope.
Direct Submission
Submitted (G5-OCT-2001) Genoscope - Centre National de Sequencage :
BP 191 91005 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Partially Hind III digested and size-selected S. mansoni cercarial
DNA was ligated into Hind III digested pBeloBAC 11 vector and used
to transform E. coli DH10B. The complete library contains 23808
clones from 4 independent sizing-ligation-transformations. Average
insert size ranges from 70-127 kb and genome coverage is 7.9-fold.
Location/Qualifiers
  1..660
    /organism="Schistosoma mansoni"
    /strain="Puerto-Rican"
    /db_xref="taxon:6183"
    /clone="023CP04"
    /clone_lib="SmBAC1"
    /note="end : T3"

BASE COUNT      139 a   16 c   9 g   446 t   50 others
ORIGIN

Query Match      5.6%   Score 96;   DB 17;   Length 660;
Best Local Similarity 46.5%;   Pred. No. 1.3e-06;
Matches 225;   Conservative 32;   Mismatches 225;   Indels 2;   Gaps 1;

QY  15  ATAACATGACAAATATTATTTCATGAGTTTATTTCGATAGCATGCACAAATATTA 74
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  615  AAAAAAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 556

QY  75  ATATATCATGTTTAAATACATGTTTTCCTTATATCACAACGACATTTTAAATTCAGACAT 134
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  555  AAAAAAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 496

QY  135  TGGTTTAAATCAATCTAA--TCTCTTATATCACAACGACATTCGCGAAATTCAGG 192
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  495  AATWAAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 436

QY  193  TAAAGAGAGAAATTAAGAGATGAGAGATGAGAGATTTCTATGGAAAGAGAGAGAGAA 252
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  435  ATATWTATATATTAATAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 376

QY  253  CATGTAGTGCACAAATAAGAGATATGATATATTTATCAGAGGTGGTGAAGATT 312
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  375  GAAAAAATAATATAAAAAATGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 316

QY  313  ATTTAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 372
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  315  KAAAGGTATTCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 256

QY  373  GAATTGTGTTAAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  255  AAAAAAATAATATAAAAAATGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 196

QY  433  TGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  195  AAAAAAATAATATAAAAAATGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 136

QY  493  AAAAA 496
    |||
DB  135  AAAAA 132

RESULT 11

```

```

AQ325799/c
LOCUS      AQ325799                982 bp    DNA             linear      GSS      05-JAN-1999
DEFINITION nBXD0021B14r CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION  nBXD0021B14r, DNA sequence.
VERSION    AQ325799
KEYWORDS   GSS.
SOURCE     Oryza sativa.
ORGANISM   Oryza sativa.

REFERENCE
AUTHORS    Wing,R.A. and Dean,R.A.
TITLE      A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL    unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAACAAGCTATGACCATG
            Class: BAC ends
            High quality sequence start: 4
            High quality sequence stop: 123.
            Location/Qualifiers
              1..982
                /organism="Oryza sativa"
                /strain="Japonica"
                /cultivar="Nipponbare"
                /db_xref="taxon:4530"
                /clone="nBXD0021B14r"
                /clone_lib="CUGI Rice BAC Library"
                /tissue_type="Leaf"
                /lab_host="E. coli DH10B"
                /note="Vector: pReloBAC11; Site:1: HindIII; Site:2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocytledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT  141 a   69 c   43 g   674 t   55 others
ORIGIN
Query Match      5.4%; Score 91; DB 17; Length 982;
Best Local Similarity 46.9%; Pred. No. 7.7e-06;
Matches 229; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

QY  9  TATTGATACATGACAATATTTATTTCATGACGTTTATTGGATACATGACACA 58
DB  903  TATGGAAATATAGAAAAGATTATTAATAAATAATTTAATTAATTAATAT 128
QY  69  ATATTATATATGCTCTTAATACATGTTTCTTCTTAATACATGATTTAAATTC 128
DB  843  TGTATTATATATATAA-TATATATTTTATATAAA-ATATATATTAATTAATTAAT 244
QY  129  AGACATTTGTTTAAATCAATCTTAATCTTTATATACACAGCATTCAGGAAAATC 188

```

```

DB  783  AATATNTATTCGTAATTAATATATATATATTTTAAATATTTTAAATATATATAAATAA 724
QY  189  CAGGTAAAGAGAGAAATAAGATGAGAGTAGAGAGATTTCTTATGGAAGAAAGAGAG 248
DB  723  AAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 664
QY  249  AGAATCATGTAGGTGAACAAAATAAAGAGATATGATGATATATTTTATGAGAGGTGGTGA 308
DB  663  AANAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 604
QY  309  GATTATTTTAGCAGGAGGAGAGAAATAGAAAAGAAATGACATGCTGAATCTGAAGA 368
DB  603  AAAANATAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 544
QY  369  AGATGAATTG-TTTAAATCATGACAGAGAGAGAAAGACACTCCATGGCTAAAGTCTCGTAA 428
DB  543  ANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 464
QY  429  AAGATGAAAGAAACAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
DB  483  ANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 424
QY  483  TGGCAAAA 496
DB  423  AAAAAAAA 416

RESULT 12
A:514657
LOCUS
DEFINITION AL514657 LT_NFL006.PL2 Homo sapiens cDNA clone CLOBB022E07.3
prime, mRNA sequence.
ACCESSION  AL514657
VERSION    AL514657.1 GI:12778151
KEYWORDS   EST.
SOURCE     Human.
ORGANISM   Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 529)
AUTHORS    Li,W.B., Gruber,C., Jesses,J. and Polayes,D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
            Location/Qualifiers
              1..529
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CLOBB0022E09"
                /clone_lib="LFL_NFL006.PL2"
                /tissue_type="placenta"
                /note="Vector: pCMVSPORT 6; Site:1: NotI; 1st strand cDNA
                was primed with a NotI-oligo(dT) primer. Five prime end
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                vector. Library was normalized. Library was constructed by
                Life Technologies. Contact : Feng Liang Life Technologies,
                a division of Invitrogen 9800 Medical Center Drive
                Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                Email : fliang@lifetech.com URL :
                http://fulllength.invitrogen.com"

BASE COUNT  364 a   20 c   9 g   89 t   47 others
ORIGIN
Query Match      5.2%; Score 89; DB 9; Length 529;
Best Local Similarity 43.9%; Pred. No. 2.1e-05;
Matches 206; Conservative 35; Mismatches 228; Indels 0; Gaps 0;

QY  28  TATTATTATTTCATGAGTTTATTTCGATGAGTACAAATATATATATACAGTGT 87

```


Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoR digestion of Drosophila DNA and provided by the RBCP from the isogenic strain Y2; cn bw sp. the same strain used for the BOP's p1 and EST libraries. A more detailed description of the library and how to order individual RAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR29B23"
 /clone_lib="RPCI-98"
 /note="end : 77"
 BASE COUNT 419 a 91 c 60 g 299 t 232 others
 ORIGIN

Query Match 5.1%; Score 86.6; DB 17; Length 1101;
 Best Local Similarity 35.8%; Prod. No. 4c-05;
 Matches 215; Conservative 96; Mismatches 270; Indels 4; Gaps 2;

```

27 5 TTTTATGGATACATGACAAATATTATTATTCATGAGTCTTTTATTCATGACGATG 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 487 TTATWAATWAAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 546

27 65 ACAATATTATATATCATGCTTTATATACAGTGTTCCTCTTAATAATACATGCTTTAA 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 ATAACTTAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 606

27 125 AATCAGACATTTGTTTAAATCAATCTATCTCTTATATCAACAGACATTCACGCGAA 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 ATAAAAAATATTTTATWAAATTTTAAATATTTTAAATATTTTAAATATTTTAAAT 666

27 185 AATTCAGTAAAGAGAGAAATATGAAGATGAGATAGAGAGATTTCTATGSAAGAA 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 667 MWATATTAATAATTTTAAATWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 726

27 245 AGAGAGACATGTAGTGAACAAATAAGAGATGATGATATTTTATGAGAGCTGG 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 TAAAWAAATAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 786

27 305 TGAAGATTATTTAGGAGGGGAGAGAGAAATAGAAAGAAATGACATGGTGAATCTG 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 787 AWAATAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 846

27 365 AAGAGATGAATTTGTTTAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 847 WAAWAAWAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 903

27 425 AAAGAGATGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 904 AAAGAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 963

27 485 CTATTCGCAAAATTTCTGTCGCGCAAAATCTACTATTTGCTCCAGGTTATTTTGG 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 964 TTWATTTWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 1022

27 545 GTTTCGAGTCAAAAGTTATTTCTTACATATATACCTGAAATATA 589
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1023 TTTAAWAAWAAATATATATWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 1067

```

RESULT 15

AZ193158/c 764 bp DNA linear GSS 30-AUG-2000
 LOCUS AZ193158
 DEFINITION SP_1022_B2_E10_SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate-1022 Col-20 Row=J, DNA sequence.
 ACCESSION AZ193158
 VERSION AZ193158.1 GI:8736337
 KEYWORDS GSS.

Strongylocentrotus purpuratus.
 Strongylocentrotus purpuratus
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 Echinodea; Euechinoides; Echinacea; Echinoida;
 Strongylocentrotidae; Strongylocentrotus.
 1 (bases 1 to 764)
 Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,I.R.,
 Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,
 G.A., Eftensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
 Hood,L.
 A sea urchin genome project: Sequence scan, virtual map, and
 additional resources
 Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
 20402566
 Contact: Cameron, RA, Davidson, EH, Hood, L
 Division of Biology 156-29
 California Institute of Technology
 Pasadena California 91125, USA
 Tel: (626) 395-8421
 Fax: (626) 793-3047
 Email: acameron@caltech.edu
 Plate: 1022 row: J column: 20
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 764.
 Location/Qualifiers
 : 764
 /organism="Strongylocentrotus purpuratus"
 /db_xref="taxon:7668"
 /clone="Plate-1022 Col=20 Row=J"
 /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
 /note="organ: sperm; Vector: BAC3.5; BAC Clones in E-Coli DH108"
 BASE COUNT 173 a 56 c 48 g 469 t 18 others
 ORIGIN
 Query Match 5.1%; Score 86.4; DB 17; Length 764;
 Best Local Similarity 47.6%; Prod. No. 5e-05;
 Matches 234; Conservative 0; Mismatches 258; Indels 0; Gaps 0;
 QY 5 TTTTATGGATACATGACAAATATTATTATTCATGAGTCTTTTATTCGATGATG 64
 DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 QY 728 TATTTAATATAGATATATAGATTTATTTATTTATTTATTTATTTATTTATTTATTT 669
 QY 65 ACAATATTATATCATGCTTTATATACAGTGTTCCTCTTAATAATACATGCTTTAA 124
 DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 669 AATTAATATTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 609
 QY 125 AATCAGACATTTGTTTAAATCAATCTATCTCTTATATCAACAGACATTTGACG 184
 DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 608 TTATAATTAATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 549
 QY 185 AATTCAGTAAAGAGAGAAATTAAGATGAGAGATGAGAGATTTCTATGGAAGAA 244
 DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 548 ATATAAATTAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 489
 QY 245 AGAGAGACATGTAGTGAACAAATAAGAGATATGATGATATTTATGAGAGCTGG 304
 DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 488 TAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 429
 QY 305 TGAAGATTATTTAGGAGGGGAGAGAGAAATAGAAAAAGAAAAATGACATGGA 364
 DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 428 AAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 369
 QY 365 AAGNAGATGATTTGTTTAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
 DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 368 AAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 309
 QY 425 AAGAGATCAAAAGAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484
 DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 308 AAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 249

QY 485 CTATTGCCAAA 496
| III:
bb 248 AAAAAAAAAA 257

Search completed: December 4, 2002, 09:21:06
JOB time : 2072 secs

GenCore version 5.1.3
Copyright (C) 1993 - 2002 Computer Ltd.

OM protein - protein search, using sw mode!

Run on: December 4, 2002, 07:18:40 : Search time 71 seconds
(without alignments)
228.966 Million cell updates/sec

Title: US-09-733-685-2

Perfect score: 623

Sequence: 1 MNLSLKEERVENGRKSDG.....QGLDLMCRPEDSVSL 122

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 13325620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_01002:*

1:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1980.DAT*
2:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1981.DAT*
3:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1982.DAT*
4:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1983.DAT*
5:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1984.DAT*
6:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1985.DAT*
7:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1986.DAT*
8:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1987.DAT*
9:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1988.DAT*
10:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1989.DAT*
11:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1990.DAT*
12:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1991.DAT*
13:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1992.DAT*
14:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1993.DAT*
15:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1994.DAT*
16:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1995.DAT*
17:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1996.DAT*
18:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1997.DAT*
19:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1998.DAT*
20:	/SID22/qcqdta/geneseq/geneseq-emb1/AA1999.DAT*
21:	/SID22/qcqdta/geneseq/geneseq-emb1/AA2000.DAT*
22:	/SID22/qcqdta/geneseq/geneseq-emb1/AA2001.DAT*
23:	/SID22/qcqdta/geneseq/geneseq-emb1/AA2002.DAT*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80.5	12.9	265	ABR00653	Chlamydia pneumoni
2	80.5	12.9	866	AA866448	Amino acid sequenc
3	80.5	12.9	968	AA778946	Polycystic kidney
4	80.5	12.9	968	AA866450	Amino acid sequenc
5	80	12.8	753	AAW91455	N. meningitidis st
6	79.5	12.8	2001	AA820062	Arabidopsis thalia
7	79.5	12.8	2001	AA820063	Arabidopsis thalia
8	78	12.5	245	AA859879	Arabidopsis thalia
9	78	12.5	250	AA859878	Arabidopsis thalia
10	78	12.5	252	AA859877	Arabidopsis thalia

11	77	12.4	685	22	ABG14755	Novel human diagno
12	77	12.4	966	23	AB807819	Mouse polycystic k
13	77	12.4	966	23	AAW51861	Murine polycystic
14	76.5	12.3	230	21	AAQ04695	Arabidopsis thalia
15	76.5	12.3	241	21	AAQ04694	Arabidopsis thalia
16	76.5	12.3	249	21	AAQ04693	Arabidopsis thalia
17	76.5	12.3	278	21	AAQ46863	Arabidopsis thalia
18	76	12.2	605	22	ABG05038	Novel human diagno
19	75	12.0	1843	22	ABG14961	Novel human diagno
20	74	11.9	141	21	AAQ46550	Arabidopsis thalia
21	74	11.9	808	23	AAQ21528	Arabidopsis thalia
22	73.5	11.8	254	22	ABG04191	Novel human diagno
23	73.5	11.8	611	22	AB861703	Drosophila melanoq
24	73.5	11.8	2059	23	ABR25711	Streptococcus poly
25	73	11.7	837	22	ABG00837	Novel human diagno
26	73	11.7	837	22	ABG12133	Novel human diagno
27	72.5	11.6	666	23	ABP26300	Streptococcus poly
28	72	11.6	255	21	AAQ60429	Arabidopsis thalia
29	72	11.6	260	21	AAQ60427	Arabidopsis thalia
30	72	11.6	262	21	AAQ60427	Arabidopsis thalia
31	71.5	11.5	167	22	AB865525	Drosophila melanoq
32	71.5	11.5	650	22	ABG04247	Novel human diagno
33	71.5	11.5	662	20	AA734751	C. pneumoniae prot
34	71	11.4	217	19	AAW80748	Human mitochondria
35	71	11.4	217	22	AAU27779	Human full-length
36	71	11.4	217	22	AAQ64577	Human GrpE Protein
37	71	11.4	215	22	ABG17196	Novel human diagno
38	71	11.4	2143	22	ABG01716	Novel human diagno
39	70.5	11.3	376	21	AAQ06245	Arabidopsis thalia
40	70.5	11.3	386	21	AAQ06244	Arabidopsis thalia
41	70.5	11.3	423	23	ABP28260	Streptococcus poly
42	70.5	11.3	551	22	AB893244	Human protein sequ
43	70.5	11.3	556	22	ABG03854	Novel human diagno
44	70.5	11.3	850	17	AA891308	Helicobacter pylor
45	70.5	11.3	859	15	AA853268	120-126 kilodalton

ALIGNMENTS

RESULT 1
ABR00653
ID: ABR00653 standard; Protein; 265 AA.
AC: ABR00653;
XX
XX
DT: 29-JUL-2002 (first entry)
XX
DE: Chlamydia pneumoniae cp6745 protein, SEQ ID NO:255.
XX

KW: Chlamydial infection; antigen; immunogen; vaccine; diagnosis;
KW: human respiratory disease; cardiovascular disease; atherosclerosis;
KW: coronary artery disease; carotid artery stenosis; myocardial infarction;
KW: cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW: strain CWL029.
XX
XX
CS: Chlamydia pneumoniae.
XX
PR: W0200202606-A2.
XX
PU: 10-CAN-2002.
XX
XX
PF: 03-JUL-2001; 2001WO-1801445.
XX
XX
PR: 03-JUL-2000; 2000GB-0016363.
PR: 11-JUL-2000; 2000GB-0017047.
PR: 21-JUL-2000; 2000GB-0017983.
PR: 07-AUG-2000; 2000GB-0019368.
PR: 18-AUG-2000; 2000GB-0020440.
PR: 10-SEP-2000; 2000GB-0022583.
PR: 10-NOV-2000; 2000GB-0027549.
PR: 22-DEC-2000; 2000GB-0031706.
XX

(CHIR) CHIRON SPA.
 Ratti G, Grandi G;
 WPI: 2002-154726/20.
 N-PSDB; ABL91311.
 Novel Chlamydia pneumoniae protein useful in the manufacture of a
 medicament for treatment or prevention of infection due to Chlamydia.
 preferably Chlamydia pneumoniae, and for diagnostic purposes
 Claim 1: Page 152: 364pp; English.
 Sequences ABB90926-ABS90715 represent novel proteins from Chlamydia
 pneumoniae (strain CW.029), and AB291184-AB291373 represent DNA encoding
 them. The proteins are predicted to be immunogenic and may therefore be
 useful in vaccine production and for diagnostic purposes. Chlamydia
 pneumoniae is a common cause of respiratory disease in humans, and is
 also involved in the development of cardiovascular diseases such as
 atherosclerosis, coronary artery disease, carotid artery stenosis,
 myocardial infarction, cerebrovascular disease, aortic aneurysm,
 claudication and stroke. The proteins and nucleic acids of the invention
 may be used in vaccines and pharmaceutical compositions for the
 prevention or treatment of chlamydial infections, particularly Chlamydia
 pneumoniae infections. The proteins may also be used in the detection of
 Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
 DNA probe assay or blotting techniques for determining Chlamydia
 pneumoniae gene expression. The present sequence represents a
 specifically claimed Chlamydia pneumoniae protein of the invention.

Sequence 265 AA;

Query Match 12.9%; Score 80.5; DB 23; Length 265;
 Best Local Similarity 27.2%; Pred. No. 1.8;
 Matches 40; Conservative 19; Mismatches 37; Indels 51; Gaps 7;
 12 EDNCKSGNRGK-----PTEVVRITVEE---VDEFFKILR-RVHVRTVAKVNGG 61
 104 QEDLGRVIGRTPEDPFVDITTEITVGLLPDLSELSTVALOGVRSRLIYVRSVK----- 159
 62 VAEGELPKKKRRKSNLGLRSLD-----CNGVR-----DGEF 94
 160 -----PMIQDLAIWGLRSDAQLNFVRLANGVQNHVPTKVKLYLAKNLADWDCEI 213
 95 DEINRVGCGCLGCLDNCPEPDSVLS 121
 214 SEERGGQKALGJD-----PKTESLSJ 236

RESUIT 2
 AAB68448
 ID AAB68448 standard; protein: 866 AA.
 AAB68448;
 23-JUL-2001 (first entry)

Amino acid sequence of an internal fragment of human PKD2.
 PKD2: ADPKD; autosomal dominant polycystic kidney disease;
 cyst formation.
 Homo sapiens.
 Key Location/Qualifiers
 FT Misc-difference 384
 FT Misc-difference 395
 FT Misc-difference 426
 FT Misc-difference 432
 FT Misc-difference 432

FT Misc-difference 547 /note= "unspecified residue"
 FT Misc-difference 586 /note= "unspecified residue"
 FT Misc-difference 587 /note= "unspecified residue"
 FT Misc-difference 588 /note= "unspecified residue"
 FT Misc-difference 849 /note= "unspecified residue"
 FT Misc-difference 849 /note= "unspecified residue"

US6228591-B1.
 06-MAY-2001.
 30-MAY-1999; 99US-0385752.
 23-MAY-1996; 96US-0651999.
 (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 Semlo S, Kochizuki T;
 WPI: 2001-342670/36.

Detecting a mutation in the PKD2 gene, useful for diagnosing autosomal
 dominant polycystic kidney disease, comprises determining the
 difference between the polynucleotide sample and the reference wild
 type PKD2 gene sequence
 Disclosure: Columns 11-18; 34pp; English.

The present sequence represents a fragment of a human PKD2 protein.
 The specification describes a method for detecting the presence of a
 mutation in the PKD2 gene. The method comprises comparing and
 determining the difference between the polynucleotide sample and the
 reference wild type PKD2 gene sequence, where the differences are
 mutations comprising one or more deletion, insertion, point or
 rearrangement mutations. The method is useful for detecting mutation
 in the PKD2 gene and for diagnosing ADPKD (autosomal dominant polycystic
 kidney disease). The method may also be used to determine whether
 persons in the population at large have ADPKD, for identifying persons
 at risk in developing the disease, e.g. relatives of the person with
 ADPKD, as well as for confirming diagnosis of ADPKD. The method may
 further be used for diagnosing ADPKD before clinical manifestations of
 the disease, i.e. formation of cysts.

Sequence 866 AA;

Query Match 12.9%; Score 80.5; DB 22; Length 866;
 Best Local Similarity 28.9%; Pred. No. 8;
 Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;
 6 EERVEEDNGKSGNRGKPSIEVVRITVEEVEFFKILRRV----HVATRTVAKVNGVA 63
 710 EEDDEDGSHSSRRRGSISSGV-----SYEEFQVLRVRVDRMEHSIGSVSKIDAVIV 762
 64 EGELPSK-KKKRSQNLGLRSLDNGV----RGEFDEINRVCGQ----- 103
 763 KLEMERAKLRREVLG--RLDQ--GVAEDRLGROSEIHRQOMERLVREELERWESDDA 818
 194 -----SLG--LDLNCKPEPS 117
 819 ASQISHSLGIPVGLNGOPRPRS 840
 RESUIT 3
 AAY78946
 ID AAY78946 standard; Protein: 968 AA.
 AAY78946;
 05-JUN-2000 (first entry)

X E Polycystic kidney disease PKD2 amino acid sequence.
 X W Polycystic kidney disease; PKD2; ADPKD; treatment; detection;
 X W Autosomal dominant polycystic kidney disease.
 X S Homo sapiens.
 X S US6031083-A.
 X D 29-FEB-2000.
 X F 23-MAY-1996; 96US-0651999.
 X R 23-MAY-1996; 96US-0651999.
 X R (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 X I Somlo S, Mochizuki T;
 X I WPI: 2000-205227/18.
 X R N-PSDB: AAZ95245.
 X X New PKD2 nucleic acid for diagnosis and treatment of autosomal dominant polycystic kidney disease associated with mutations in the PKD2 gene -
 T T Claim 1: Fig 5: 35pp; English.
 X S This sequence represents the human polycystic kidney disease (PKD2) protein sequence. The PKD2 gene is associated with Autosomal dominant Polycystic Kidney Disease (ADPKD). The PKD2 gene and protein can be used for the diagnosis and treatment of ADPKD caused by defects in the PKD2 gene. ADPKD can be diagnosed by detecting the presence of a mutated PKD2 gene in nucleic acid sequences from an individual. Diagnosis of high risk patients prior to the appearance of clinical manifestations is also possible. Detection is carried out by sequencing, restriction enzyme digestion analysis or hybridization of probes to wild type or mutated PKD2 gene or by detecting expression of the gene product i.e. the mRNA or protein product. ADPKD caused by a mutated PKD2 gene can be treated by delivery and expression of a functional PKD2 gene into a sufficient number of cells of the subject to treat the disease.

XX OS Homo sapiens.
 XX PN US6228591-B1.
 XX PD 28-MAY-2001.
 XX PF 30-AUG-1999; 96US-0385752.
 XX PR 23-MAY-1996; 96US-0651999.
 XX PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 XX PI Somlo S, Mochizuki T;
 XX DR WPI: 2001-342670/35.
 XX DR N-PSDB: AAF85440.
 XX PT Detecting a mutation in the PKD2 gene, useful for diagnosing autosomal dominant polycystic kidney disease, comprises determining the difference between the polynucleotide sample and the reference wild type PKD2 gene sequence -
 XX PS Disclosure: Fig 5A-G: 34pp; English.
 XX CC The present sequence represents a human PKD2 protein. The specification describes a method for detecting the presence of a mutation in the PKD2 gene. The method comprises comparing and determining the difference between the polynucleotide sample and the reference wild type PKD2 gene sequence, where the differences are mutations comprising one or more deletion, insertion, point or rearrangement mutations. The method is useful for detecting mutation in the PKD2 gene and for diagnosing ADPKD (autosomal dominant polycystic kidney disease). The method may also be used to determine whether persons in the population at large have ADPKD, for identifying persons at risk in developing the disease, e.g. relatives of the person with ADPKD, as well as for confirming diagnosis of ADPKD. The method may further be used for diagnosing ADPKD before clinical manifestations of the disease, i.e. formation of cysts.
 XX SQ Sequence 968 AA;
 Query Match 12.9%; Score 80.5; DB 22; Length 968;
 Best Local Similarity 28.9%; Pred. No. 9.2;
 Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;
 QY 8 EERVEEDNGKSDGNRGKPSDEVVTVTVEEVDFFKILRRV-----HVAITVAKVNGVA 63
 DB 813 EEDDEDSDGHSRRRGSTISGGV-----SYEEFQVLVRRVDRMERHSIGSVSKIDAVIV 865
 QY 64 EGEPLSK-KRKRSONLGRNSLDCNGV----RDGEFDEINRVGLQ----- 103
 DB 866 KLEIMBRALKRREVLS--RLLD--GVAEDRLGRDSEIHREQMERLVREELERWESDDA 921
 QY 104 -----GLG--LDLNCKPEPDS 117
 DB 922 ASQISHGLGTPVGLNGQPRPS 943
 RESULT 5
 AAW93495
 ID AAW93495 standard; Protein: 753 AA.
 XX AC AAW93495;
 XX AC AAW93495;
 XX DT 11-JUN-1999 (first entry)
 XX DE N. meningitidis strain M990 LbpB protein.
 XX DE LbpB: lactoferrin binding protein; vaccine; neisserial disease;
 KW meningitis; diagnosis; treatment.
 XX CS Neisseria meningitidis.
 XX CS

PN W09909176-AL.
 XX 25-FEB-1999.
 XX 10-AUG-1998: 98WC-EP05117.
 XX 05-FEB-1998: 98CB-CC02544.
 PR 15-AUG-1997: 97GB-C017423.
 XX (UYUT-) RIJKSUNIV UTRECHT.
 PA (TECH-) TECHNOLOGY FOUND TECHNOLOGIESTICHTING ST.
 PI Pettersson-Fernholm AM, Icmasssen JPM;
 XX WPI: 1999-190165/16.
 DR N-PSDB: AAX23322.
 XX New lactoferrin-binding protein B polynucleotides - obtained from
 PT Neisseria meningitidis. Used to develop products for the diagnosis,
 PT prevention and treatment of neisserial disease, e.g. meningitis
 XX Claim 13; Page 102-105: 116pp; English.
 PS This invention describes novel lactoferrin-binding protein B (LbPB;
 CC strains of Neisseria meningitidis. The products of this invention can
 CC be used for vaccinating humans against neisserial disease e.g.
 CC meningitis. Antibodies raised against the proteins of the invention
 CC can be used for diagnosing or treating neisserial disease in humans.
 CC The LbPB polypeptides can also be used for identifying compounds which
 CC inhibit the polypeptides.
 XX Sequence 753 AA:
 SQ
 Query Match 12.8%; Score 80; DB 20; Length 753;
 Best Local Similarity 27.8%; Pred. No. 7.6;
 Matches 25; Conservative 18; Mismatches 33; Indels 14; Gaps 3;
 JY 6 KKEERVEEDNGKSDGNRGPSTVEVPTVTEEEVDFFKILRRVHVATRTVAKVNGVAEG 65
 DB 496 EDETSDEONGDEATATETEEVDEAEDEVEE-----PEENSPAGNNG-SGS 545
 QY 66 ELPKSKRKRSQNLGRNSLDGNSVRDGEFD 95
 DB 546 ILPALEASKGRDI---DLFLKGIPTAETD 571
 RESULT 6
 AAB20062
 ID AAB20062 standard; Protein: 2001 AA.
 XX AAB20062:
 XX 23-APR-2001 (first entry)
 XX Arabidopsis thaliana silencing gene-uncoded protein.
 XX Gene silencing; silencing gene; MOM.
 XX Arabidopsis thaliana.
 XX Key Location/Qualifiers
 FH Region 177..350
 FT /note= "internal repeat region"
 FT Peptide 362..367
 FT /note= "nuclear localization motif"
 FT Binding-site 450..467
 FT /note= "ATP/GTP-binding motif"
 FT Region: 479..719
 FT /note= "region of homology to Arpase/helicase
 family SW12/SNF proteins"
 FT Misc-difference 705
 FT /note= "Lys in ecotype Columbia"
 FT Peptide 832..838

FT Peptide /note= "nuclear localization motif"
 FT 858..862
 FT /note= "nuclear localization motif"
 FT Domain 995..1015
 FT /note= "predicted membrane-spanning domain"
 FT Misc-difference 1219
 FT /note= "Asp in ecotype Columbia"
 FT Region 1462..1672
 FT /note= "internal repeat region"
 FT Region 1848..1894
 FT /note= "internal repeat region"
 FT Domain 1899..1941
 FT /note= "actin-binding domain"
 XX W020010601 A2.
 PN 04-JAN-2001.
 PD 21-JUN-2000: 2000WC-EF05761.
 FF 21-JUN-1999: 99GB-C014623.
 PR (NOVS) NOVARTIS AG.
 XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX Hazou Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
 XX WPI: 2001-137952/14.
 DR N-PSDB: AAA89353, AAA89354.
 XX Novel gene encoding a protein that controls gene silencing in
 PT particular silencing of plant genes
 PT Claim 1; Page 32-39: 48pp; English.
 XX The present sequence is that of Arabidopsis thaliana protein MOM,
 CC which controls gene silencing, particularly gene silencing. It
 CC is encoded by a novel silencer gene (see AAA89353) of A. thaliana
 CC ecotype Zurich. The protein exhibits partial similarity with
 CC Arpase/helicase proteins of the SW12/SNF2 family. The sequence
 CC differs at amino acids 705 and 1219 compared with the protein
 CC (see AAB20063) encoded by the silencer gene of wild-type A.
 CC thaliana ecotype Columbia. Gene silencing is useful as a molecular
 CC tool for regulating gene expression.
 XX Sequence 2001 AA:
 SQ
 Query Match 12.8%; Score 79.5; DB 22; Length 2001;
 Best Local Similarity 21.9%; Pred. No. 30;
 Matches 25; Conservative 24; Mismatches 34; Indels 31; Gaps 4;
 QY 5 LKKEERV-----EEDNGKSDGNRGPSTVEVPTVEEVEDEFFK 43
 DB 1 MKKDEKGLTGRTIYRSLSAASPASVEQETPGLRSSRQTPSTIKVTPASATKSF--- 57
 QY 44 ILRRVHVATRTVAKVNGVAEGEJPSKKRRSQ---NLGRNSLDC-NGVRDGE 93
 DB 55 ---RLAPSPASVSKKSGGIVKNSTPSSILRSNRGKTEVSLQSSKSGSDNSIRKGD 108
 RESULT 7
 AAB20063
 ID AAB20063 standard; Protein: 2001 AA.
 XX AAB20063:
 XX 23-APR-2001 (first entry)
 XX Arabidopsis thaliana silencing gene-uncoded protein.
 XX Gene silencing; silencing gene; MOM.
 XX Arabidopsis thaliana.

[illegible]

ID	AA051861 standard; Protein; 966 AA.	KW	Protein identification; signal transduction pathway; metabolic pathway;
XX	AA051861;	KW	hybridisation assay; genetic mapping; gene expression control; promoter;
XX	29-JAN-2002 (first entry)	XX	termination sequence.
XX	Murine polycystic kidney disease protein 2.	OS	Arabidopsis thaliana.
DE	Mouse; TLCC-2; TRP-like calcium channel; membrane excitability;	PN	EPI033405-A2
XX	nociception; neuroprotective; antiparkinsonian; cytostatic;	PD	06-SEP-2000.
XX	hypotensive; antidepressant; analgesic; anticonvulsant; tranquiliser;	XX	25-FEB-2000; 2000EP-0301439.
XX	Parkinson's disease; Huntington's disease; multiple sclerosis;	XX	25-FEB-1999; 99US-0121825.
KW	Gilles de la Tourette's syndrome; autonomic function disorder; cancer;	PR	05-MAR-1999; 99US-0123180.
KW	neuroleptic; gene therapy; Alzheimer's disease; CNS disorder.	PR	09-MAR-1999; 99US-0123548.
XX	Mus musculus.	PR	23-MAR-1999; 99US-0125788.
OS	WO200177331-A1.	PR	25-MAR-1999; 99US-0126264.
PN	18-OCT-2001.	PR	29-MAR-1999; 99US-0126785.
XX	06-APR-2001; 2001WO-US11442.	PR	01-APR-1999; 99US-0127452.
XX	07-APR-2000; 2000US-C544797.	PR	06-APR-1999; 99US-0128234.
XX	(MILL-) MILLENIUM PHARM INC.	PR	09-APR-1999; 99US-0128714.
XX	Curtis RAJ, Silos-Santiago I;	PR	16-APR-1999; 99US-0129845.
XX	WPI; 2002-01013/01.	PR	19-APR-1999; 99US-0130077.
XX	Novel isolated human transient receptor potential-like calcium channel	PR	21-APR-1999; 99US-0130449.
PT	protein-2 useful for treating Alzheimer's disease, depression, amnesia,	PR	23-APR-1999; 99US-0130891.
PT	pain disorder, and cancer	PR	28-APR-1999; 99US-0131449.
XX	Example 1; Fig 7; 148pp; English.	PR	30-APR-1999; 99US-0132048.
XX	The present invention relates to the protein and coding sequences of	PR	04-MAY-1999; 99US-0132484.
CC	human transient receptor potential (TRP)-like calcium channel protein-2	PR	05-MAY-1999; 99US-0132485.
CC	(TLCC-2). The sequences can be used in the treatment of TLCC-2 related	PR	06-MAY-1999; 99US-0132486.
CC	disorders, including central nervous system disorders such as	PR	06-MAY-1999; 99US-0132487.
CC	Alzheimer's, Parkinson's and Huntington's diseases, multiple sclerosis,	PR	11-MAY-1999; 99US-0132963.
CC	Gilles de la Tourette's syndrome, autonomic function disorders, learning	PR	14-MAY-1999; 99US-0134218.
CC	or memory disorders, pain disorders and disorders of cellular	PR	14-MAY-1999; 99US-0134221.
CC	proliferation, including cancer. The present sequence is the murine	PR	14-MAY-1999; 99US-0134370.
CC	polycystic kidney disease protein 2 which was described in the	PR	18-MAY-1999; 99US-0134768.
CC	exemplification of the invention.	PR	19-MAY-1999; 99US-0134941.
XX	Sequence 966 AA:	PR	20-MAY-1999; 99US-0135124.
SQ	Query Match 12.4%; Score 77; DB 23; Length 966;	PR	21-MAY-1999; 99US-0135353.
	Best Local Similarity 29.1%; Pred. No. 2;	PR	24-MAY-1999; 99US-0135629.
	Matches 30; Conservative 17; Mismatches 44; Indels 12; Gaps 3;	PR	25-MAY-1999; 99US-0136021.
QY	8 EERVEEDNKSQGNRGKPEVTVTEBEVDFEKKILRV----RVATRYAKVNGGVA 63	PR	27-MAY-1999; 99US-0136392.
	: : : : : : : : : : : : : :	PR	28-MAY-1999; 99US-0136782.
Db	811 EEEDEDSGHSSRRRSISGV-----SYEEFQVLVRVDMHSHSISVSKTDATV 963	PR	01-JUN-1999; 99US-0137528.
	: : : : : : : : : : : : : :	PR	03-JUN-1999; 99US-0137528.
QY	64 EGELPSK-KKKRSQNLGLNSLDCNGVRGGEDEINRVQLGL 105	PR	04-JUN-1999; 99US-0137502.
	: : : : : : : : : : : : : :	PR	07-JUN-1999; 99US-0137724.
Db	864 KLEIMPRAKLKRREVLGLDGVAEADRLCRDSEIHREQMERL 906	PR	08-JUN-1999; 99US-0138094.
	: : : : : : : : : : : : : :	PR	10-JUN-1999; 99US-0138540.
RESULT 14		PR	14-JUN-1999; 99US-0139119.
AA004695		PR	16-JUN-1999; 99US-0139452.
XX	AA004695 standard; Protein; 230 AA.	PR	16-JUN-1999; 99US-0139453.
XX	AA004695;	PR	17-JUN-1999; 99US-0139492.
XX	17-OCT-2000 (first entry)	PR	18-JUN-1999; 99US-0139454.
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 840.	PR	18-JUN-1999; 99US-0139455.
DE		PR	18-JUN-1999; 99US-0139456.
XX		PR	18-JUN-1999; 99US-0139457.
		PR	18-JUN-1999; 99US-0139458.
		PR	18-JUN-1999; 99US-0139459.
		PR	18-JUN-1999; 99US-0139460.
		PR	18-JUN-1999; 99US-0139461.
		PR	18-JUN-1999; 99US-0139462.
		PR	18-JUN-1999; 99US-0139463.
		PR	18-JUN-1999; 99US-0139750.
		PR	18-JUN-1999; 99US-0139763.
		PR	21-JUN-1999; 99US-0139817.
		PR	22-JUN-1999; 99US-0139899.
		PR	23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140595.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141267.
PR 01-JUL-1999; 99US-0141642.
PR 01-JUL-1999; 99US-0142554.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142677.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145152.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148342.
PR 13-AUG-1999; 99US-0148505.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149724.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149910.
PR 25-AUG-1999; 99US-0150556.
PR 26-AUG-1999; 99US-0150684.
PR 27-AUG-1999; 99US-0151064.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PK 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157751.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159264.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 16-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.38; Score 76.5; DB 21; Length 230;
Best Local Similarity 29.88; Pred. No. 4;
Matches 25; Conservative 13; Mismatches 29; Indels 17; Gaps 4;
QY 2 NSLKKKEERVEEDNGKSDGNRGKPESTEVYRTVTEEVDEFFKILRRVHVAVRTVAKVNGG 61
DB 122 NNRLLNDSDKEE---IDNNQKDD---RSVYEELE-----RLH---RKIDELKGG 154
QY 52 VAEGELPSKKRKRSONGLRNSLD 85
DD 165 RGSDDPRKSNERRKKNLNFKKKRD 188

RESULT 15
AA04594
ID AA04594 standard; Protein: 241 AA.
XX AA04594;
XX AC

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 819.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX	OS	Arab:Gopsis thaliana.	PR	29-JUN-1999;	99US-0140991.
XX	PN	EPI033405-A2.	PR	30-JUN-1999;	99US-0141267.
XX	XX	06-SEP-2000.	PR	01-JUL-1999;	99US-0141842.
XX	XX		PR	02-JUL-1999;	99US-0142154.
XX	XX		PR	03-JUL-1999;	99US-0142055.
XX	XX		PR	04-JUL-1999;	99US-0142390.
XX	XX		PR	05-JUL-1999;	99US-0142520.
XX	XX		PR	06-JUL-1999;	99US-0142977.
XX	XX		PR	07-JUL-1999;	99US-0143542.
XX	XX		PR	08-JUL-1999;	99US-0143624.
XX	XX		PR	09-JUL-1999;	99US-0144005.
XX	XX		PR	10-JUL-1999;	99US-0144085.
XX	XX		PR	11-JUL-1999;	99US-0144086.
XX	XX		PR	12-JUL-1999;	99US-0144325.
XX	XX		PR	13-JUL-1999;	99US-0144331.
XX	XX		PR	14-JUL-1999;	99US-0144332.
XX	XX		PR	15-JUL-1999;	99US-0144333.
XX	XX		PR	16-JUL-1999;	99US-0144334.
XX	XX		PR	17-JUL-1999;	99US-0144335.
XX	XX		PR	18-JUL-1999;	99US-0144352.
XX	XX		PR	19-JUL-1999;	99US-0144632.
XX	XX		PR	20-JUL-1999;	99US-0144632.
XX	XX		PR	21-JUL-1999;	99US-0144884.
XX	XX		PR	22-JUL-1999;	99US-0144884.
XX	XX		PR	23-JUL-1999;	99US-0145056.
XX	XX		PR	24-JUL-1999;	99US-0145088.
XX	XX		PR	25-JUL-1999;	99US-0145085.
XX	XX		PR	26-JUL-1999;	99US-0145087.
XX	XX		PR	27-JUL-1999;	99US-0145069.
XX	XX		PR	28-JUL-1999;	99US-0145192.
XX	XX		PR	29-JUL-1999;	99US-0145145.
XX	XX		PR	30-JUL-1999;	99US-0145218.
XX	XX		PR	31-JUL-1999;	99US-0145224.
XX	XX		PR	01-AUG-1999;	99US-0145276.
XX	XX		PR	02-AUG-1999;	99US-0145913.
XX	XX		PR	03-AUG-1999;	99US-0145918.
XX	XX		PR	04-AUG-1999;	99US-0145919.
XX	XX		PR	05-AUG-1999;	99US-0145951.
XX	XX		PR	06-AUG-1999;	99US-0146386.
XX	XX		PR	07-AUG-1999;	99US-0146388.
XX	XX		PR	08-AUG-1999;	99US-0146389.
XX	XX		PR	09-AUG-1999;	99US-0147038.
XX	XX		PR	10-AUG-1999;	99US-0147204.
XX	XX		PR	11-AUG-1999;	99US-0147302.
XX	XX		PR	12-AUG-1999;	99US-0147192.
XX	XX		PR	13-AUG-1999;	99US-0147260.
XX	XX		PR	14-AUG-1999;	99US-0147303.
XX	XX		PR	15-AUG-1999;	99US-0147416.
XX	XX		PR	16-AUG-1999;	99US-0147493.
XX	XX		PR	17-AUG-1999;	99US-0147935.
XX	XX		PR	18-AUG-1999;	99US-0148171.
XX	XX		PR	19-AUG-1999;	99US-0148319.
XX	XX		PR	20-AUG-1999;	99US-0148341.
XX	XX		PR	21-AUG-1999;	99US-0148565.
XX	XX		PR	22-AUG-1999;	99US-0148684.
XX	XX		PR	23-AUG-1999;	99US-0149368.
XX	XX		PR	24-AUG-1999;	99US-0149175.
XX	XX		PR	25-AUG-1999;	99US-0149426.
XX	XX		PR	26-AUG-1999;	99US-0149722.
XX	XX		PR	27-AUG-1999;	99US-0149723.
XX	XX		PR	28-AUG-1999;	99US-0149929.
XX	XX		PR	29-AUG-1999;	99US-0149902.
XX	XX		PR	30-AUG-1999;	99US-0149930.
XX	XX		PR	31-AUG-1999;	99US-0150566.
XX	XX		PR	01-SEP-1999;	99US-0150884.
XX	XX		PR	02-SEP-1999;	99US-0151065.
XX	XX		PR	03-SEP-1999;	99US-0151066.
XX	XX		PR	04-SEP-1999;	99US-0151080.
XX	XX		PR	05-SEP-1999;	99US-0151303.
XX	XX		PR	06-SEP-1999;	99US-0151438.
XX	XX		PR	07-SEP-1999;	99US-0151930.
XX	XX		PR	08-SEP-1999;	99US-0152363.
XX	XX		PR	09-SEP-1999;	99US-0153070.

```
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157885.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159437.
PR 14-OCT-1999; 99US-0159538.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.3%; Score 76.5; DB 21; Length 241;
Best Local Similarity 29.8%; Pred. No. 4.3;
Matches 25; Conservative 13; Mismatches 29; Indels 17; Gaps 4;
QY 2 NNSLKKEEERVEEDNGKSDGNRGKPSPEVTVTTEEEVDEFFKILRAVHVAVRTVAKVNGG 61
DB 133 NNRINDDDSKEE----TDNNRQKDD----RSVTYEMLRE-----RLH---RKIDELKGG 175
QY 62 VARGELFSKKRKPSQNLGLKNSLD 85
DB 176 RGSDDPRPSNERKKKLNLPKKRRD 199
```

Search completed: December 4, 2002, 09:22:29
Job time : 79 secs

GenCore version: 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2002, 09:22:38 ; Search time 27 seconds
(without alignments)
132,948 Million cell updates/sec

Title: US-09-733-685-2
Perfect score: 623
Sequence: 1 MNNSLKKEERVEEDNKGKSG.....QQLGGLDLNCKIEPDVSLSL 122

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents-AA: *
1: /cgn2.6/prodata/1/aa/5A_COMB.pep: *
2: /cgn2.6/prodata/1/aa/5B_COMB.pep: *
3: /cgn2.6/prodata/1/aa/5A_COMB.pep: *
4: /cgn2.6/prodata/1/aa/5B_COMB.pep: *
5: /cgn2.6/prodata/1/aa/PC1US_COMB.pep: *
6: /cgn2.6/prodata/1/aa/backfilest.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DR	ID	Description
1	80.5	12.9	866	3	US-08-651-999A-1	Sequence 1, Appl
2	80.5	12.9	866	4	US-09-385-752-1	Sequence 1, Appl
3	80.5	12.9	968	3	US-08-651-999A-7	Sequence 7, Appl
4	80.5	12.9	968	4	US-09-385-752-7	Sequence 7, Appl
5	71	11.4	217	3	US-08-971-158-1	Sequence 1, Appl
6	71	11.4	217	4	US-05-416-488-1	Sequence 1, Appl
7	70.5	11.3	859	1	US-08-053-614-2	Sequence 1, Appl
8	70.5	11.3	859	1	US-08-316-397B-2	Sequence 2, Appl
9	70.5	11.3	859	2	US-09-034-306-2	Sequence 2, Appl
10	70.5	11.3	859	4	US-09-259-437-2	Sequence 2, Appl
11	70.5	11.3	859	5	PCT-US93-09782-2	Sequence 2, Appl
12	70.5	11.3	1181	1	US-08-053-614-4	Sequence 4, Appl
13	70.5	11.3	1181	1	US-08-316-397B-4	Sequence 4, Appl
14	70.5	11.3	1181	2	US-09-034-306-4	Sequence 4, Appl
15	70.5	11.3	1181	4	US-09-259-437-4	Sequence 4, Appl
16	70.5	11.3	1181	5	PCT-US93-09782-4	Sequence 4, Appl
17	69	11.1	591	1	US-08-145-999A-21	Sequence 21, Appl
18	69	11.1	591	2	US-08-451-747-21	Sequence 21, Appl
19	69	11.1	591	3	US-09-134-852-21	Sequence 21, Appl
20	66.5	10.7	349	4	US-08-955-916C-2	Sequence 2, Appl
21	66.5	10.7	424	4	US-09-661-711A-8	Sequence 8, Appl
22	66.5	10.7	455	4	US-09-661-711A-4	Sequence 4, Appl
23	66.5	10.7	455	4	US-08-661-711A-26	Sequence 26, Appl
24	66.5	10.7	542	1	US-08-701-380-2	Sequence 2, Appl
25	66.5	10.7	542	3	US-09-032-365A-13	Sequence 13, Appl
26	66.5	10.7	551	4	US-08-661-711A-2	Sequence 2, Appl
27	66	10.6	251	2	US-08-592-214A-12	Sequence 12, Appl

28 56 10.6 251 3 US-08-659-188-12 Sequence 12, Appl
29 66 10.6 251 3 US-08-655-227-12 Sequence 12, Appl
30 66 10.6 251 3 US-08-655-241-12 Sequence 12, Appl
31 66 10.6 251 3 US-09-149-976-12 Sequence 12, Appl
32 66 10.6 251 4 US-09-398-326-12 Sequence 12, Appl
33 66 10.6 552 4 US-09-661-711A-10 Sequence 10, Appl
34 65.5 10.5 349 4 US-08-697-766A-2 Sequence 2, Appl
35 65.5 10.5 1024 4 US-09-562-737-82 Sequence 82, Appl
36 65.5 10.5 1170 4 US-09-749-588-2 Sequence 2, Appl
37 65.5 10.5 15281 2 US-08-471-119A-2 Sequence 281, Appl
38 65 10.4 180 4 US-09-615-192A-281 Sequence 4, Appl
39 65 10.4 196 3 US-07-998-289B-4 Sequence 51, Appl
40 64.5 10.4 148 1 US-08-166-195A-51 Sequence 51, Appl
41 64.5 10.4 148 2 US-08-436-883B-51 Sequence 51, Appl
42 64.5 10.4 599 1 US-08-752-238-3 Sequence 3, Appl
43 64 10.3 599 3 US-09-085-603B-3 Sequence 3, Appl
44 64 10.3 599 3 US-09-031-897-7 Sequence 7, Appl
45 64 10.3 599 3

ALIGNMENTS

RESULT 1
US-08-651-999A-1
; Sequence 1, Application US/08651999A
; Patent No. 6031086
; GENERAL INFORMATION:
; APPLICANT: Stefan Somlo and Toshio Mochizuki
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSCHILD & EHENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
; MEDIUM TYPE: DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651.999A
; FILING DATE: MAY 23, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/395
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TW 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 866 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: peptide
; HYPOTHEetical: NO
; FRAGMENT TYPE: internal fragment
US-08-651-999A-1

Query Match 12.9%; Score 80.5; DB 3; Length 866;
Best Local Similarity 28.9%; Pred. No. 0.59;
Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;

QY 8 EERVEEDNKGSDGRGKSTEVRTVTEEEVDEEFKILRRV----HVATRTVAKVNGVA 63

DB 710 EDDDEDSGHSRRRGSSGV-----SYEEFQVLVRVDRMEHSIGSVSKIDAVI 752

QY 64 EGELPSK-KRKRSONGLNRSLDCNGV-----RDGEFDEINRVGLQ----- 103

DB 763 KLEINERAKLRREVLG--RLID--GVAEDERLGRDSEIHREQMERLVREELERWESDUA 818

QY 104 -----GLG--LDLNCKRPEPS 117

DB 819 ASQISHGLGTPVGLNGQPRPS 840

RESULT 2

US-09-385-752-1
: Sequence 1, Application US/09385752
: Patent No. 6228591
: GENERAL INFORMATION:
: APPLICANT: Stefan Somlo and Toshio Mochizuki
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
: TITLE OF INVENTION: AND USES THEREOF
: NUMBER OF SEQUENCES: 12

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
: STREET: 90 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: U.S.A.
: ZIP: 10016

: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
: MEDIUM TYPE: DISKETTE
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/385,752
: FILING DATE: MAY 23, 1996

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/651,995
: FILING DATE: MAY 23, 1996

: ATTORNEY/AGENT INFORMATION:
: NAME: ELIZABETH A. BOGOSIAN
: REGISTRATION NUMBER: 39,911
: REFERENCE/DOCKET NUMBER: 96700/395
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 697-5595
: TELEFAX: (212) 286-0854 or 286-0082
: TELEX: TWX 710-581-4766

: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 866 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE:
: DESCRIPTION: peptide
: HYPOTHEICAL: NO
: FRAGMENT TYPE: internal fragment

US-09-385-752-1
Query Match 12.98; Score 80.5; DB 4; Length 866;
Best Local Similarity 28.98; Pred. No. 0.59;
Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;

QY 8 EERVEEDNCKSDGNKGPSTEVVRITVEEVEFFKILRRV-----HYVATRVAKVNGVA 63

DB 710 EDDDEDSGHSRRRGSSGV-----SYEEFQVLVRVDRMEHSIGSVSKIDAVI 752

QY 64 EGELPSK-KRKRSONGLNRSLDCNGV-----RDGEFDEINRVGLQ----- 103

DB 763 KLEINERAKLRREVLG--RLID--GVAEDERLGRDSEIHREQMERLVREELERWESDUA 818

QY 104 -----GLG--LDLNCKRPEPS 117

DB 819 ASQISHGLGTPVGLNGQPRPS 840

RESULT 3

US-08-651-999A-7
: Sequence 7, Application US/08651999A
: Patent No. 6031089
: GENERAL INFORMATION:
: APPLICANT: Stefan Somlo and Toshio Mochizuki
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
: TITLE OF INVENTION: AND USES THEREOF
: NUMBER OF SEQUENCES: 12

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
: STREET: 90 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: U.S.A.
: ZIP: 10016

: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
: MEDIUM TYPE: DISKETTE
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/651,999A
: FILING DATE: MAY 23, 1996

: ATTORNEY/AGENT INFORMATION:
: NAME: ELIZABETH A. BOGOSIAN
: REGISTRATION NUMBER: 39,911
: REFERENCE/DOCKET NUMBER: 96700/395

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 697-5595
: TELEFAX: (212) 286-0854 or 286-0082
: TELEX: TWX 710-581-4766

: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 988 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE:
: DESCRIPTION: peptide
: HYPOTHEICAL: NO

US-08-651-999A-7
Query Match 12.98; Score 80.5; DB 3; Length 968;
Best Local Similarity 28.98; Pred. No. 0.59;
Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;

QY 8 EERVEEDNCKSDGNKGPSTEVVRITVEEVEFFKILRRV-----HYVATRVAKVNGVA 63

DB 813 EDDDEDSGHSRRRGSSGV-----SYEEFQVLVRVDRMEHSIGSVSKIDAVI 865

QY 64 EGELPSK-KRKRSONGLNRSLDCNGV-----RDGEFDEINRVGLQ----- 103

DB 866 KLEINERAKLRREVLG--RLID--GVAEDERLGRDSEIHREQMERLVREELERWESDUA 921

QY 104 -----GLG--LDLNCKRPEPS 117

DB 922 ASQISHGLGTPVGLNGQPRPS 943

RESULT 4

US-09-385-752-7
: Sequence 7, Application US/09385752
: Patent No. 6228591
: GENERAL INFORMATION:
: APPLICANT: Stefan Somlo and Toshio Mochizuki
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
: TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: AMSTER, FCTUSTEIN & EHRENSTEIN
 STREET: 90 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: U.S.A.
 ZIP: 10016

COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE

MEDIUM TYPE: DISKETTE

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/385.752

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/651.999

FILING DATE: MAY 23, 1996

ATTORNEY/AGENT INFORMATION:

NAME: ELIZABETH A. BOGOSIAN

REGISTRATION NUMBER: 39,911

REFERENCE/DOCKET NUMBER: 96700/395

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 697-5995

TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 968 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE:

DESCRIPTION: peptide

HYPOTHEICAL: NO

US-09-385-752-7

Query Match: 12.9% Score 90.5; DB 4; Length 968;

Best Local Similarity 28.9%; Pred. No. 0.697

Matches 41; Conservative 19; Mismatches 19; Indels 43; Gaps 8;

QY 8 ERVEEDNGKSGNKGKPESTVEVTVTEEEVEFEKTLRRV---HVATRTVAKYNGGVA 63

DB 813 EEDDDSDSHSSRRSGSISSGV-----SYEEQVLYRVVVRMEHSIGSLVSKIDAVIV 865

QY 64 EGELPSK-KRKRSQNLGRNSLDGNGV-----RDGPFDEINRVGLGLD-----NCKPEP 115

DB 856 KLEIMERAKLRREVIG--RLLD--GVAEDEKGPQSEHRECMERLYRELELRNESDVA 921

QY 104 -----GLG--LDLCKPEFS 117

DB 922 ASQISHGLGTPVGLNGQAPRS 943

US-08-971-158-1

Sequence 1, Application US/08971.58

Patent No. 6010879

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Goli, Surya K.

TITLE OF INVENTION: HUMAN MITOCHONDRIAL CHAPERONE PROTEIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/971.158

FILING DATE: Herewith

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/824,875

FILING DATE: March 26, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0245-1 CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 217 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLNOT08

CLONE: 1844121

US-08-971-158-1

Query Match: 11.4% Score 71; DB 3; Length 217;

Best Local Similarity 25.6%; Pred. No. 1.2;

Matches 31; Conservative 16; Mismatches 32; Indels 42; Gaps 6;

QY 9 ERVEEDNGKSGNKGKPESTVEVTVTEEEV--DEFFKILRRVHVATRTVAKYNGGVAE 69

DB 38 QNLDEDMQCSFQKADPPATE--KTLLEKVKLEPQLA-----ETVEKYKPALADTE 86

QY 67 LPSKKRKRSQNLGRNSLDGNGVDRGDFEINRVGLGLD-----NCKPEP 115

DB 87 ---NLRORSOKL-----VEEAKLYGIOAFCKDLLEADVLEKATCCVPKE 128

QY 116 D 116

DB 129 E 129

US-09-416-488-1

Sequence 1, Application US/094.6488

Patent No. 6435915

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Goli, Surya K.

TITLE OF INVENTION: HUMAN MITOCHONDRIAL CHAPERONE PROTEIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/416.488

FILING DATE: 12-Oct-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

Db 614 LEKSLKREHLEKDVAKNLESCKGNKKMEAKAQAANSQKDEI--FALINKEANRARAIA 671

QY 57 KVGNGVAEGELPSKKRKSQNLGRN-----SLDNGVRDGEFDEINRV-----GLQGL 105

Db 672 YAQ-----NLGKIKRELSQKLENIKDKDFSKSDFGFKNGKNKQFSKAETLKALGKS 725

QY 106 GLDLNCKPE 114

Db 726 VKDLGINPE 734

RESULT 9

US-09-034-306-2

Sequence 2, Application US/09054306

Patent No. 5876943

GENERAL INFORMATION:

APPLICANT: COVER, TIMOTHY L.

APPLICANT: BLASER, MARTIN J.

APPLICANT: HARRY KLEANIHOU

APPLICANT: TUMMURU, MURALI K.R.

TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING

TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P.C.

STREET: 127 Peachtree Street, Suite 1200

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/034.306

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/316.397

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: SPRATT, GWENDOLYN D.

REGISTRATION NUMBER: 36,016

REFERENCE/DOCKET NUMBER: 2200.030

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404/688-0770

TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 859 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-034-306-2

Query Match 11.3% Score 70.5; DB 2; Length 859;

Best Local Similarity 24.0%; Pred. No. 9;

Matches 31; Conservative 21; Mismatches 54; Indels 23; Gaps 5;

QY 1 MNSLKKERVEEDNGKS-----DGNRKPSDEVVTVTEEVDEFFKILRRVHVATRIVA 56

Db 614 LEKSLKREHLEKDVAKNLESCKGNKKMEAKAQAANSQKDEI--FALINKEANRARAIA 671

QY 57 KVGNGVAEGELPSKKRKSQNLGRN-----SLDNGVRDGEFDEINRV-----GLQGL 105

Db 672 YAQ-----NLGKIKRELSQKLENIKDKDFSKSDFGFKNGKNKQFSKAETLKALGKS 725

QY 106 GLDLNCKPE 114

Db 726 VKDLGINPE 734

RESULT 10

US-09-259-437-2

Sequence 2, Application US/09259437

Patent No. 6153390

GENERAL INFORMATION:

APPLICANT: COVER, TIMOTHY L.

APPLICANT: BLASER, MARTIN J.

APPLICANT: HARRY KLEANIHOU

APPLICANT: TUMMURU, MURALI K.R.

TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING

TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P.C.

STREET: 127 Peachtree Street, Suite 1200

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/259.437

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/316.397

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: SPRATT, GWENDOLYN D.

REGISTRATION NUMBER: 36,016

REFERENCE/DOCKET NUMBER: 2200.030

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404/688-0770

TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 859 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-259-437-2

Query Match 11.3% Score 70.5; DB 4; Length 859;

Best Local Similarity 24.0%; Pred. No. 9;

Matches 31; Conservative 21; Mismatches 54; Indels 23; Gaps 5;

QY 1 MNSLKKERVEEDNGKS-----DGNRKPSDEVVTVTEEVDEFFKILRRVHVATRIVA 56

Db 614 LEKSLKREHLEKDVAKNLESCKGNKKMEAKAQAANSQKDEI--FALINKEANRARAIA 671

QY 57 KVGNGVAEGELPSKKRKSQNLGRN-----SLDNGVRDGEFDEINRV-----GLQGL 105

Db 672 YAQ-----NLGKIKRELSQKLENIKDKDFSKSDFGFKNGKNKQFSKAETLKALGKS 725

QY 106 GLDLNCKPE 114

Db 726 VKDLGINPE 734

RESULT 11

PCT-US93-09782-2

Sequence 2, Application PC/TUS9309782

GENERAL INFORMATION:

APPLICANT: COVER, TIMOTHY L.

APPLICANT: BLASER, MARTIN J.

APPLICANT: TUMMURU, MURALI K.R.

TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING

Matches 31; Conservative 21; Mismatches 54; Indels 23; Gaps 5;

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

DB 726 VKDLGINPE 734

Search completed: December 4, 2002, 05:26:48
Job time : 30 secs

GenCore version 5.1.3
Copyright (C) 1993 - 2002 Computer, Ltd.

QM protein - protein search, using sw model

Run on: December 4, 2002, 09:23:23 : Search time 21 seconds
(without alignments)
94,360 Million cell updates/sec

Title: US-09-733-685-2

Perfect score: 623

Sequence: 1 MNSLXKXERVEDNKRSG.....QGLGLDKNKPELVSLSL 122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 103943 seqs, 16242309 residues

Total number of hits satisfying chosen parameters: 103943

Minimum DB seq length: 0

Maximum DB seq length: 2500000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*

2: /cgn2_5/ptodata/2/pubpaa/PCI_NEW_PUB pep:*

3: /cgn2_5/ptodata/2/pubpaa/US06_NEW_PUB pep:*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*

6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*

7: /cgn2_6/ptodata/2/pubpaa/PCRTUS_PUBCOMB pep:*

8: /cgn2_5/ptodata/2/pubpaa/US06_PUBCOMB pep:*

9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*

11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*

12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*

13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*

14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	Description
1	80.5	12.9	866	10	US-09-753-008-1 Sequence 1, Appli
2	80.5	12.9	968	10	US-09-753-008-7 Sequence 7, Appli
3	77	12.4	966	10	US-09-628-466-6 Sequence 6, Appli
4	70	11.2	251	10	US-09-335-390A-30 Sequence 30, Appli
5	70	11.2	326	10	US-09-745-763-150 Sequence 150, Appl
6	70	11.2	325	10	US-09-325-301-969 Sequence 301, Appl
7	69.5	11.2	491	10	US-09-815-242-12023 Sequence 12023, A
8	67.5	10.8	240	10	US-09-401-369-342 Sequence 342, App
9	67.5	10.8	653	12	US-10-052-566-402 Sequence 402, App
10	66.5	10.7	159	10	US-09-815-242-5016 Sequence 5016, Ap
11	66.5	10.7	257	10	US-09-815-242-10681 Sequence 10681, A
12	66	10.6	155	10	US-09-925-301-1370 Sequence 1370, Ap
13	66	10.6	390	10	US-09-815-242-10683 Sequence 10683, A
14	65.5	10.5	1170	12	US-10-135-687-2 Sequence 2, Appli
15	65.5	10.5	1210	10	US-09-860-352A-2 Sequence 2, Appli
16	65	10.4	89	10	US-09-864-761-36753 Sequence 36753, A
17	64.5	10.4	162	10	US-09-864-761-33534 Sequence 33534, A
18	64.5	10.4	712	10	US-09-815-242-10930 Sequence 10930, A
19	64.5	10.4	2568	10	US-09-866-108-3 Sequence 3, Appli

Sequence 18, Appli
Sequence 5, Appli
Sequence 4, Appli
Sequence 312, App
Sequence 188, App
Sequence 3, Appli
Sequence 3, Appli
Sequence 161, App
Sequence 2, Appli
Sequence 1294, Ap
Sequence 10, Appl
Sequence 4, Appli
Sequence 2041, A
Sequence 7, Appli
Sequence 14, Appl
Sequence 24, Appl
Sequence 6, Appli
Sequence 13261, A
Sequence 4, Appli
Sequence 13765, A
Sequence 234, App
Sequence 34, Appl
Sequence 82, Appl
Sequence 70, Appl
Sequence 4, Appli
Sequence 12, Appli

ALIGNMENTS

RESULT 1
US-09-753-008-1
Sequence 1, Application US/09753008
Patent No. US20020061520A1
GENERAL INFORMATION:
APPLICANT: Stefan Som-O and Toshio Mochizuki
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
AND USES THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/753,008
FILING DATE: 02-Jan-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/385,752
FILING DATE: 1999-08-30

APPLICATION NUMBER: 08/651,999
FILING DATE: MAY 23, 1996

ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN

REGISTRATION NUMBER: 39,911

REFERENCE/DOCKET NUMBER: 96700/395

TELEPHONE: (212) 697-5995

TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 1

SEQUENCE CHARACTERISTICS:

LENGTH: 866 amino acids

TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: <Unknown>
DESCRIPTION: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: internal fragment
SEQUENCE DESCRIPTION: SEQ ID NO: 1
US-09-753-008-1

Query Match 12.9%; Score 80.5; DB 10; Length 868;
Best Local Similarity 28.9%; Pred. No. 1;
Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;
QY 8 EERVEEDNGKSGNRGKPTVEVYVITEEVEDEFFKILRRV----HVATRTVAKYNGVA 63
DB 710 EEDDEDGSHSSRRGSISSGV-----SYEEFQVLVRRVDRMEHSIGSVSKIDAVIV 762
QY 64 EGELPSK-KKKRSONLGRNSLDCNGV----RDGEFDEINRVGLQ-----103
DB 763 KLEIMRAKLRREVIG--RLLD--GVAFDERLGRDSEIHRQEMERLVREELERWESODA 818
QY 104 -----GLG--LDLNCKPEPDS 117
DB 819 ASOISHGGLGTPVGLNCPRPDS 840

RESULT 2

US-09-753-008-7
Sequence 7, Application US/09753008
Patent No. US20020061520A1
GENERAL INFORMATION:
APPLICANT: Stefan Somlo and Toshio Mochizuki
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
AND USES THEREOF

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
DISKETTE

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/753,008
FILING DATE: 02-Jan-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/385,752
FILING DATE: 1999-08-30
APPLICATION NUMBER: 08/551,999
FILING DATE: MAY 23, 1996

ATTORNEY/AGENT INFORMATION:

NAME: ELIZABETH A. ROGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/395
TELEPHONE: (212) 697-5995
TELEFAX: (212) 296-0854 or 286-0082
TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 7

SEQUENCE CHARACTERISTICS:
LENGTH: 968 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: <Unknown>
DESCRIPTION: peptide
HYPOTHETICAL: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 7
US-09-753-008-7

Query Match 12.9%; Score 80.5; DB 10; Length 968;
Best Local Similarity 28.9%; Pred. No. 1.2;
Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;
QY 8 EERVEEDNGKSGNRGKPTVEVYVITEEVEDEFFKILRRV----HVATRTVAKYNGVA 63
DB 813 EEDDEDGSHSSRRGSISSGV-----SYEEFQVLVRRVDRMEHSIGSVSKIDAVIV 865
QY 64 EGELPSK-KKKRSONLGRNSLDCNGV----RDGEFDEINRVGLQ-----103
DB 866 KLEIMRAKLRREVIG--RLLD--GVAFDERLGRDSEIHRQEMERLVREELERWESODA 921
QY 104 -----GLG--LDLNCKPEPDS 117
DB 922 ASOISHGGLGTPVGLNCPRPDS 943

RESULT 3

US-09-828-466-6
Sequence 6, Application US/09828466
Patent No. US20020035056A1
GENERAL INFORMATION:
APPLICANT: Curtis, Rory A.J.
APPLICANT: Sinos-Santiago, Immaculada
TITLE OF INVENTION: 54420, A NOVEL HUMAN CALCIUM CHANNEL
FILE REFERENCE: MNI-125CP
CURRENT APPLICATION NUMBER: US/09/828,466
CURRENT FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: US 09/544,797
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 6
LENGTH: 966
TYPE: PRT
ORGANISM: Homo sapiens
US-09-828-466-6

Query Match 12.4%; Score 77; DB 10; Length 966;
Best Local Similarity 29.1%; Pred. No. 2.8;
Matches 30; Conservative 17; Mismatches 44; Indels 12; Gaps 3;

QY 8 EERVEEDNGKSGNRGKPTVEVYVITEEVEDEFFKILRRV----HVATRTVAKYNGVA 53
DB 811 EEDDEDGSHSSRRGSISSGV-----SYEEFQVLVRRVDRMEHSIGSVSKIDAVIV 863
QY 64 EGELPSK-KKKRSONLGRNSLDCNGVVRDGEFDEINRVGLQL 105
DB 864 KLEIMRAKLRREVIG--RLLD--GVAFDERLGRDSEIHRQEMERL 906

RESULT 4

US-09-635-390A-30
Sequence 30, Application US/09035390A
Patent No. US20020076761A1
GENERAL INFORMATION:
APPLICANT: Escobedo, Jaime
Oulianjin, Hu
Garcia, Pablo
Williams, Lewis T.
Kothakota, Srinivas
TITLE OF INVENTION: Secreted Human Proteins
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916

1 PRIOR APPLICATION NUMBER: 60/066772
2 PRIOR FILING DATE: 1997-11-24
3 PRIOR APPLICATION NUMBER: 60/069335
4 PRIOR FILING DATE: 1997-12-22
5 PRIOR APPLICATION NUMBER: 60/069425
6 PRIOR FILING DATE: 1997-12-22
7 PRIOR APPLICATION NUMBER: 60/069670
8 PRIOR FILING DATE: 1997-12-17
9 PRIOR APPLICATION NUMBER: 60/068617
10 PRIOR FILING DATE: 1997-12-18
11 PRIOR APPLICATION NUMBER: 60/077450
12 PRIOR FILING DATE: 1998-03-10
13 PRIOR APPLICATION NUMBER: 60/077632
14 PRIOR FILING DATE: 1998-03-11
15 PRIOR APPLICATION NUMBER: 60/077649
16 PRIOR FILING DATE: 1998-03-11
17 PRIOR APPLICATION NUMBER: 60/078886
18 PRIOR FILING DATE: 1998-04-20
19 PRIOR APPLICATION NUMBER: 60/078939
20 PRIOR FILING DATE: 1998-03-20
21 PRIOR APPLICATION NUMBER: 60/079664
22 PRIOR FILING DATE: 1998-03-27
23 PRIOR APPLICATION NUMBER: 60/079786
24 PRIOR FILING DATE: 1998-03-27
25 PRIOR APPLICATION NUMBER: 60/080107
26 PRIOR FILING DATE: 1998-03-31
27 PRIOR APPLICATION NUMBER: 60/080194
28 PRIOR FILING DATE: 1998-03-31
29 PRIOR APPLICATION NUMBER: 60/080327
30 PRIOR FILING DATE: 1998-04-01
31 PRIOR APPLICATION NUMBER: 60/080333
32 PRIOR FILING DATE: 1998-04-01
33 PRIOR APPLICATION NUMBER: 60/081049
34 PRIOR FILING DATE: 1998-04-08
35 PRIOR APPLICATION NUMBER: 60/081070
36 PRIOR FILING DATE: 1998-04-08
37 PRIOR APPLICATION NUMBER: 60/081195
38 PRIOR FILING DATE: 1998-04-09
39 PRIOR APPLICATION NUMBER: 60/081838
40 PRIOR FILING DATE: 1998-04-15
41 PRIOR APPLICATION NUMBER: 60/082568
42 PRIOR FILING DATE: 1998-04-21
43 PRIOR APPLICATION NUMBER: 60/082568
44 PRIOR FILING DATE: 1998-04-21
45 PRIOR APPLICATION NUMBER: 60/082704
46 PRIOR FILING DATE: 1998-04-22
47 PRIOR APPLICATION NUMBER: 60/082797
48 PRIOR FILING DATE: 1998-04-22
49 PRIOR APPLICATION NUMBER: 60/083322
50 PRIOR FILING DATE: 1998-04-28
51 PRIOR APPLICATION NUMBER: 60/083495
52 PRIOR FILING DATE: 1998-04-29
53 PRIOR APPLICATION NUMBER: 60/083496
54 PRIOR FILING DATE: 1998-04-29
55 PRIOR APPLICATION NUMBER: 60/083499
56 PRIOR FILING DATE: 1998-04-29
57 PRIOR APPLICATION NUMBER: 60/083559
58 PRIOR FILING DATE: 1998-04-29
59 PRIOR APPLICATION NUMBER: 60/084366
60 PRIOR FILING DATE: 1998-05-05
61 PRIOR APPLICATION NUMBER: 60/084414
62 PRIOR FILING DATE: 1998-05-06
63 PRIOR APPLICATION NUMBER: 60/084639
64 PRIOR FILING DATE: 1998-05-07
65 PRIOR APPLICATION NUMBER: 60/084640
66 PRIOR FILING DATE: 1998-05-07
67 PRIOR APPLICATION NUMBER: 60/084643
68 PRIOR FILING DATE: 1998-05-07
69 PRIOR APPLICATION NUMBER: 60/085573
70 PRIOR FILING DATE: 1998-05-15
71 PRIOR APPLICATION NUMBER: 60/085579
72 PRIOR FILING DATE: 1998-05-15
73 PRIOR APPLICATION NUMBER: 60/085580

74 PRIOR FILING DATE: 1998-05-15
75 PRIOR APPLICATION NUMBER: 60/085882
76 PRIOR FILING DATE: 1998-05-15
77 PRIOR APPLICATION NUMBER: 60/085700
78 PRIOR FILING DATE: 1998-05-15
79 PRIOR APPLICATION NUMBER: 60/086023
80 PRIOR FILING DATE: 1998-05-18
81 PRIOR APPLICATION NUMBER: 60/086392
82 PRIOR FILING DATE: 1998-05-22
83 PRIOR APPLICATION NUMBER: 60/086486
84 PRIOR FILING DATE: 1998-05-22
85 PRIOR APPLICATION NUMBER: 60/087098
86 PRIOR FILING DATE: 1998-05-28
87 PRIOR APPLICATION NUMBER: 60/087208
88 PRIOR FILING DATE: 1998-05-28
89 PRIOR APPLICATION NUMBER: 60/087609
90 PRIOR FILING DATE: 1998-06-02
91 PRIOR APPLICATION NUMBER: 60/087759
92 PRIOR FILING DATE: 1998-06-02
93 PRIOR APPLICATION NUMBER: 60/087827
94 PRIOR FILING DATE: 1998-06-03
95 PRIOR APPLICATION NUMBER: 60/088025
96 PRIOR FILING DATE: 1998-06-04
97 PRIOR APPLICATION NUMBER: 60/088028
98 PRIOR FILING DATE: 1998-06-04
99 PRIOR APPLICATION NUMBER: 60/088029
100 PRIOR FILING DATE: 1998-06-04
101 PRIOR APPLICATION NUMBER: 60/088033
102 PRIOR FILING DATE: 1998-06-04
103 PRIOR APPLICATION NUMBER: 60/088167
104 PRIOR FILING DATE: 1998-06-05
105 PRIOR APPLICATION NUMBER: 60/088202
106 PRIOR FILING DATE: 1998-06-05
107 PRIOR APPLICATION NUMBER: 60/088212
108 PRIOR FILING DATE: 1998-06-05
109 PRIOR APPLICATION NUMBER: 60/088217
110 PRIOR FILING DATE: 1998-06-05
111 PRIOR APPLICATION NUMBER: 60/088326
112 PRIOR FILING DATE: 1998-06-04
113 PRIOR APPLICATION NUMBER: 60/088655
114 PRIOR FILING DATE: 1998-06-09
115 PRIOR APPLICATION NUMBER: 60/088722
116 PRIOR FILING DATE: 1998-06-10
117 PRIOR APPLICATION NUMBER: 60/088736
118 PRIOR FILING DATE: 1998-06-10
119 PRIOR APPLICATION NUMBER: 60/088740
120 PRIOR FILING DATE: 1998-06-10
121 PRIOR APPLICATION NUMBER: 60/088811
122 PRIOR FILING DATE: 1998-06-10
123 PRIOR APPLICATION NUMBER: 60/088824
124 PRIOR FILING DATE: 1998-06-10
125 PRIOR APPLICATION NUMBER: 60/088825
126 PRIOR FILING DATE: 1998-06-10
127 PRIOR APPLICATION NUMBER: 60/088826
128 PRIOR FILING DATE: 1998-06-10
129 PRIOR APPLICATION NUMBER: 60/088861
130 PRIOR FILING DATE: 1998-06-11
131 PRIOR APPLICATION NUMBER: 60/088863
132 PRIOR FILING DATE: 1998-06-11
133 PRIOR APPLICATION NUMBER: 60/088876
134 PRIOR FILING DATE: 1998-06-11
135 PRIOR APPLICATION NUMBER: 60/089090
136 PRIOR FILING DATE: 1998-06-12
137 PRIOR APPLICATION NUMBER: 60/089105
138 PRIOR FILING DATE: 1998-06-12
139 PRIOR APPLICATION NUMBER: 60/089512
140 PRIOR FILING DATE: 1998-06-16
141 PRIOR APPLICATION NUMBER: 60/089514
142 PRIOR FILING DATE: 1998-06-16
143 PRIOR APPLICATION NUMBER: 60/089538
144 PRIOR FILING DATE: 1998-06-17
145 PRIOR APPLICATION NUMBER: 60/089598
146 PRIOR FILING DATE: 1998-06-17

;; PRIOR APPLICATION NUMBER: 60/089053
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089058

Query Match 10.8%; Score 67.5; DB 12; Length 653;

Best Local Similarity 23.9%; Pred. No. 17; Mismatches 55; Indels 29; Gaps 4;
Matches 32; Conservative 18

Qy 1 MNLSKKEEVEEDKSGDNKPKSTEVYRTVEEVEDEFFKILRP-----VHVATR 53

Db 429 MOKSLERREKAEKEEKQTKKTETSAEKLRLKEE--KPLKKKKRKSISSSSVSSAUE 486

Qy 54 TVAKVNGVGAEGELPSKKRKRSONGLRNS-----LDNGVDCGDFEINRVGLQL 107

Db 487 SVSSSSSSSGHKRHKRKRNSRSSRRHSSKSSNQIDQNKDE----- 535

Qy 108 DLNCKPPDSVSL 121

Db 536 ---CYPVPANTSAS 546

RESULT 10

US-09-615-242-5016

;; Sequence 5016, Application US/09815242

;; Patent No. US20020061569A1

;; GENERAL INFORMATION:

;; APPLICANT: Haselbeck, Robert

;; APPLICANT: Ohlsen, Karl L.

;; APPLICANT: Zyskind, Judith W.

;; APPLICANT: Wall, Daniel

;; APPLICANT: Trawick, John D.

;; APPLICANT: Carr, Grant J.

;; APPLICANT: Yamamoto, Robert I.

;; APPLICANT: Xu, H. Howard

;; TITLE OF INVENTION: Identification of Essential Genes in

;; TITLE OF INVENTION: Prokaryotes

;; FILE REFERENCE: ELITRA.011A

;; CURRENT APPLICATION NUMBER: US/09/815,242

;; PRIOR FILING DATE: 2001-03-21

;; PRIOR APPLICATION NUMBER: 60/191,078

;; PRIOR FILING DATE: 2000-03-21

;; PRIOR APPLICATION NUMBER: 60/206,848

;; PRIOR FILING DATE: 2000-05-23

;; PRIOR APPLICATION NUMBER: 60/207,727

;; PRIOR FILING DATE: 2000-05-26

;; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR FILING DATE: 2000-10-23

;; PRIOR APPLICATION NUMBER: 60/253,625

;; PRIOR FILING DATE: 2000-11-27

;; PRIOR APPLICATION NUMBER: 60/257,931

;; PRIOR FILING DATE: 2000-12-22

;; PRIOR APPLICATION NUMBER: 60/269,308

;; PRIOR FILING DATE: 2001-02-16

;; NUMBER OF SEQ ID NOS: 1411C

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 5016

;; LENGTH: 159

;; TYPE: PRT

;; ORGANISM: Enterococcus faecalis

US-09-815-242-5015

Query Match 10.7%; Score 66.5; DB 10; Length 159;

Best Local Similarity 28.9%; Pred. No. 3.4;

Matches 26; Conservative 11; Mismatches 40; Indels 13; Gaps 3;

Qy 19 DGNRCKPSTEVYRTVEEVEDEFFKIL---RRVHVATRTVAKYNGVGAEGELPSKKRKS 75

Db 75 DGMVKPRLRVETINEEGTSVYRLVQEGKTHQVKNLMTATGKPV-----YLRKLRM 129

Qy 76 QNLGLRNSLDGNGVRDGEFDEINRVGLQL 105

Db 130 GELWLDETEL-----GEYRPLTEARLQQL 154

RESULT 11

US-09-815-242-10681

;; Sequence 10681, Application US/09815242

;; Patent No. US20020061569A1

;; GENERAL INFORMATION:

;; APPLICANT: Haselbeck, Robert

;; APPLICANT: Ohlsen, Karl L.

;; APPLICANT: Zyskind, Judith W.

;; APPLICANT: Wall, Daniel

;; APPLICANT: Trawick, John D.

;; APPLICANT: Carr, Grant J.

;; APPLICANT: Yamamoto, Robert I.

;; APPLICANT: Xu, H. Howard

;; TITLE OF INVENTION: Identification of Essential Genes in

;; TITLE OF INVENTION: Prokaryotes

;; FILE REFERENCE: ELITRA.011A

;; CURRENT APPLICATION NUMBER: US/09/815,242

;; CURRENT FILING DATE: 2001-03-21

;; PRIOR APPLICATION NUMBER: 60/191,078

;; PRIOR FILING DATE: 2000-03-21

;; PRIOR APPLICATION NUMBER: 60/206,848

;; PRIOR FILING DATE: 2000-05-23

;; PRIOR APPLICATION NUMBER: 60/207,727

;; PRIOR FILING DATE: 2000-05-26

;; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR FILING DATE: 2000-10-23

;; PRIOR APPLICATION NUMBER: 60/253,625

;; PRIOR FILING DATE: 2000-11-27

;; PRIOR APPLICATION NUMBER: 60/257,931

;; PRIOR FILING DATE: 2000-12-22

;; PRIOR APPLICATION NUMBER: 60/269,308

;; PRIOR FILING DATE: 2001-02-16

;; NUMBER OF SEQ ID NOS: 1411C

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 10681

;; LENGTH: 257

;; TYPE: PRT

;; ORGANISM: Enterococcus faecalis

US-09-815-242-10681

Query Match 10.7%; Score 66.5; DB 10; Length 257;

Best Local Similarity 28.9%; Pred. No. 6.4;

Matches 26; Conservative 11; Mismatches 40; Indels 13; Gaps 3;

Qy 19 DGNRCKPSTEVYRTVEEVEDEFFKIL---RRVHVATRTVAKYNGVGAEGELPSKKRKS 75

Db 170 DGMVKPRLRVETINEEGTSVYRLVQEGKTHQVKNLMTATGKPV-----YLRKLRM 224

Qy 76 QNLGLRNSLDGNGVRDGEFDEINRVGLQL 105

Db 225 GELWLDETEL-----GEYRPLTEARLQQL 249

RESULT 12

US-09-925-301-1370

;; Sequence 1370, Application US/09925301

;; Patent No. US20020052308A1

;; GENERAL INFORMATION:

;; APPLICANT: Rosen, et al.

;; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

;; FILE REFERENCE: PA106

;; CURRENT APPLICATION NUMBER: US/09/925,301

;; CURRENT FILING DATE: 2001-08-10

;; PRIOR APPLICATION NUMBER: PCT/US00/03882

;; PRIOR FILING DATE: 2000-03-08

;; PRIOR APPLICATION NUMBER: 60/124,270

;; PRIOR FILING DATE: 1999-03-12

;; NUMBER OF SEQ ID NOS: 1694

;; SOFTWARE: PatentIn Ver. 2.0

;; SEQ ID NO 1370

;; LENGTH: 155

;; TYPE: PRT

2
3
4
5
6
7
8

Search completed: December 4, 2002, 09:27:28
Job time : 22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compaq Inc.

OM protein - protein search, using sw model

Run on: December 4, 2002, 09:21:19 ; Search time 44 seconds
(without alignments)
256,555 Million cells updates/sec

Title: US-09-733-685-2

Perfect score: 623

Sequence: 1 MNNSLKKEERVEDNGKSDG.....GGGGLGGLNKKPKRHSVSL 122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 6.5

Searched: 283224 seqs, 9634422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY-ES

Result No.	Score	Query Match	Length	DB ID	Description
1	80.5	12.9	265	2 A86548	hypothetical prote
2	80.5	12.9	285	2 H72075	hypothetical prote
3	80.5	12.9	604	2 G02640	polycystic kidney
4	80.5	12.9	2254	2 D86215	protein T5022.14
5	78.5	12.6	245	2 C81592	hypothetical prote
6	76	12.2	1007	2 H81670	conserved hypote
7	74	11.9	141	2 T00928	hypothetical prote
8	74	11.9	307	2 G69211	conserved hypote
9	74	11.9	891	2 G96436	hypothetical prote
10	73.5	11.8	389	2 C86450	P1421.8 protein -
11	73	11.7	107	2 C81232	50S ribosomal prot
12	73	11.7	350	2 T48522	hypothetical prote
13	73	11.7	2149	2 T47555	hypothetical prote
14	72.5	11.6	448	2 S63961	hypothetical prote
15	72.5	11.6	484	2 H50140	transcription, reg
16	72.5	11.6	804	2 T32864	hypothetical prote
17	72	11.6	132	2 T50108	yeast ARG2 prot
18	71.5	11.5	335	2 D83142	hypothetical prote
19	71.5	11.5	543	2 S53817	thermosome beta sh
20	71.5	11.5	2275	2 T33123	hypothetical prote
21	71	11.4	467	2 D66485	protein: F28J9.13
22	71	11.4	862	2 T01141	hypothetical prote
23	70.5	11.3	302	1 TPCHRC	treponem T. nard
24	70.5	11.3	314	2 T29363	hypothetical prote
25	70.5	11.3	775	1 WMBE19	ribonucleoside-dip
26	70.5	11.3	2297	2 AB2494	hypothetical prote
27	70	11.2	227	2 S57480	mcllydopterin-guan
28	70	11.2	853	2 T23697	hypothetical prote
29	69.5	11.2	255	2 D71404	hypothetical prote

ALIGNMENTS

RESULT 1

A86548

hypothetical protein CPJ0461 [imported] - Chlamydia pneumoniae (strain J138)

C:Species: Chlamydia pneumoniae; Chlamydia pneumoniae

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001

C:Accession: A86548

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: A86548

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-265 <STO>

A:Cross-references: GB:BAQ00006; NID:9878831; PID:BAA38667.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: CPJ0461

Query Match 12.9%; Score 80.5; DB 2; Length 255;

Best Local Similarity 27.2%; Pred. NC. 3-9;

Matches 40; Conservative 19; Mismatches 37; Indels 51; Gaps 7;

QY 12 EFQDKSDGNKSGK-----PSIEVVRVTVEE---VDEFFKILR-RVYVATRTVAKVNS 62

DB 104 QFDGSKVGRTPPEPFDVDTPTIEVQLLPDEELSTVDEALQGVRSNLTAYRSVEK----- 159

QY 62 VAEGELPSKKRKRSONGLRNSLD-----CNGVR-----DGEF 94

DB 160 -----PMIIDLAVGEGIRDSABLNFVRLANGVQNHYPHTVKLYLAKNLADVWDEK 213

QY 95 DFNRVGTGGLGLO:NCKPPEPSVLS 121

DB 214 SEERKGLRALGID-----PRTESISLT 236

RESULT 2

H72075

hypothetical protein - Chlamydia pneumoniae (strain CWL029)

C:Species: Chlamydia pneumoniae; Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000

C:Accession: H72075

R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, J.; Grimwood,

Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Accession: H72075

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-265 <ARN>

A:Cross-references: GB:AE001630; GB:AE001630; NID:94376740; PID:RAD18603.1; PID:9437

A:Experimental source: strain CWL029

[illegible]

```
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: G59211
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, R.; Dubois, C.; Aldredge, T.
O: Liu, D.; Shadafara, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani,
K.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A59000; MUID:98037514; PMID:9371463
A:Accession: G59211
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 5307 <MTE>
A:Cross-references: GB:AE000860; GB:AE000666; NID:g2621915; PIDN:AAB85335.1; P.D:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH838

Query Match 11.9% Score 74: DB 2: Length 307:
Best Local Similarity 26.9% Pred. No. 18:
Matches 52: Conservative 13: Mismatches 38: Indels 36: Gaps 4:

QY 5 LKKEERVEEDNGKSDGN-----RGKDPSTEVRTVTEEEVDEFFKI----- 44
DB 3 LKQ:ERKTSQSPSDT.LRVKGLVIMGKPIAEAIRSVKGESEETWEISEKESHGDRYD 52
QY 45 ---LRRVIVATRVAKVNGVGVGELPSKKRKRKSNGLRNSLDCNVRDGEFEINRV 100
DB 63 VEKLRIIRIAQRSLRNQGSTIE----RLKRKRKSLLGKIRELE-----DEXSL 103

RESULT 9
G59636
Hypothetical protein Flp17.17 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G59636
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, G.; Aton
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.A.; Devar,
ansen, D.F.; Hughes, B.; Huizuar, L.
Nature 406, 836-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
A.; Liu, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Mervia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A8641; MUID:21016719; PMID:11130712
A:Accession: G59636
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1:691 <STO>
A:Cross-references: GB:AE005173; NID:g2443891; PIDN:AAB71484.1; GSPDB:GN06141
C:Genetics:
A:Gene: Flp17.17
A:Map position: 1

Query Match 11.9% Score 74: DB 2: Length 891:
Best Local Similarity 28.6% Pred. No. 60:
Matches 22: Conservative 16: Mismatches 33: Indels 6: Gaps 2:

QY 2 NNSLKKSERVEEDNGKSDGNRGKDPSTEVRTVTEEEVDEFFKILRRVHVATRVAKVNGG 61
DB 784 SNGTAQEK--EDKSGEKNVKNVPETRSKTSPTSRATETMSKTAKKPTVASRMAQKNKF 841
QY 62 VAEGELPSKKRKRKSNQL 78
DB 842 EKEEEM---KKRIENL 854

RESULT 10
G59640
Fl4M2.1a protein - Arabidopsis thaliana
```

```
A:Molecule type: DNA
A:Residues: 1-1007 <TEI>
A:Cross-references: GB:AE002342; GB:AE002160; NID:g719076; PIDN:AAF39550.1; PID:g719077
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001
C:Accession: T00928; H64850
R:Kounis, S.B.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
A:Description: The Arabidopsis thaliana chromosome 11 BAC T24P15 genomic sequence.
A:Reference number: 214212
A:Accession: T00928
A:Status: translated from GB/EMBL/DBRC
A:Molecule type: DNA
A:Residues: 1-141 <ROU>
A:Cross-references: EMBL:AC002561; NID:g2673901; PID:g2673910
A:Experimental source: Cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, J.; Talbot, J.
ous, D.; Kleiman, W.C.; White, G.; Eskin, G.; Eschen, J.A.; Salzberg, S.B.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: H64850
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-141 <STO>
A:Cross-references: GB:AE002093; NID:g2673920; PIDN:AAB88344.1; GSPDB:GN06139
C:Genetics:
A:Gene: T24P15.10; At2g42190
A:Map position: 2
C:Superfamily: Arabidopsis thaliana hypothetical protein: T0KI7.140

Query Match 11.9% Score 74: DB 2: Length 141:
Best Local Similarity 19.7% Pred. No. 7.8:
Matches 26: Conservative 18: Mismatches 40: Indels 48: Gaps 4:

QY 6 KKEERVEEDNGKSD-----GNRGKDPSTEVRTVTEEEVDEFFKILRRVHVATRVAKV 58
DB 10 KQNASAREDRGGSGDEEK-PAYRRRGKPKMKDDFEEDDEEDDEEVEKMEVEDKCEID 69
QY 59 NGGVAEGELPSKKR-----KSNGLGRNS----- 83
DB 70 DGSVTSKDKERKRKMSNGSNTDLTTEENGLGSKPNTDSTSTSTGTFQNGSRKRKSKPR 129
QY 84 -----LDCNGV 89
DB 130 RAAEAVVECNV 141

RESULT 8
G59211
conserved hypothetical protein MTH838 - Methanobacterium thermoautotrophicum (strain Del
```

C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov 2001
C:Accession: C86460
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, C.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.R.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Klaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MIMD:21016719; PMID:11130712
A:Accession: C86460
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <STO>
A:Cross-references: GB:AE005172; NID:q9665104; P1EN:AAF57295.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
Query Match 11.8%; Score 73.5; DB 2; Length 189;
Best Local Similarity 28.3%; Pred. No. 27;
Matches 30; Conservative 17; Mismatches 38; Indels 21; Gaps 6;
QY 7 KEERVEDNCKSGNR-GKPSSTVVR-TVTEVEDFFKILRRVHVATVAKVNGGVAE 64
DB 82 KEESTKPTAQSDENOGNPLIEKVGSTIDAEESD-----KKMEDNESKEVNISSGONE 135
QY 65 GELPSKK-----KKRSNGLNRLSDNGVRD-----GEFDEIN 98
DB 136 GEDDSKETNDVVAQKEVEN-GSKEVTPCDSQKDEANAGSGEKN 150
RESULT 11
C81232
50S ribosomal protein L24 NMR0153 [imported] - Neisseria meningitidis (strain MC58 sero-
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Feb-2001
C:Accession: C81232; D82004
A:Title: Complete genome sequence of Neisseria meningitidis serogroup A strain MC58.
A:Reference number: A81006; MIMD:20175755; PMID:10710307
A:Accession: C81232
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-107 <NET>
A:Cross-references: GB:AE002373; GB:AE002098; NID:q7225359; P1EN:AAF50611.1; P1D:q722537
A:Experimental source: serogroup A, strain MC58
A:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.P.; Morel,
Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mangall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z3491.
A:Reference number: A81775; MIMD:20225556; PMID:10761919
A:Accession: D82004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-107 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:q7378778; PID:CA893433.1; PID:q737882
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: rplX; NMR0153; NMA0118
C:Superfamily: Escherichia coli ribosomal protein L24
Query Match 11.7%; Score 73; DB 2; Length 107;
Best Local Similarity 24.3%; Pred. No. 7.1;
Matches 25; Conservative 22; Mismatches 34; Indels 22; Gaps 4;

QY 3 NSLKKERVEDNCKSGNRGKPSSTVVR-TVTEVEEDFFKILRRVHVATVAKVNGGVA 62
DB 2 NKIKGRVVVVIAGKKGKGG-----QVAVLGDKVVEGVNVVVKR-HOKPNPMRJTEGI 56
QY 63 AEGELPSKKKKRQNLGRNSLDCNGVR--DGEFDEINRVGLQ 103
DB 57 TKEMP-----LDISNIALNPETNKADRVGK 84
RESULT 12
C46822
hypothetical protein F18022.260 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48622
R:Boyan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.;
Submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24493
A:Accession: T48622
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-350 <HEV>
A:Cross-references: EMBL:DB
A:Experimental source: cultivar Columbia; BAC clone F18022
C:Genetics:
A:Map position: 5
A:Introns: 50/3; 113/1; 275/3
A:Note: F18022.280
Query Match 11.7%; Score 73; DB 2; Length 350;
Best Local Similarity 28.4%; Pred. No. 26;
Matches 27; Conservative 13; Mismatches 29; Indels 26; Gaps 4;
QY 7 KEERVEDNCKSGNRGKPSSTVVR-TVTEVEEDFFKILRR 47
DB 220 KQEKTTIDSESSGVGGPSTPTTITQVVPVWISVDEDFADDDSKMVLSEYAEGLN 275
QY 48 VHVATR-----TVAKVNGGVA--EGELPSKKKKRS 75
DB 250 ICEASSSTSDKIAKVGNGVSVIEDNLMKKIEAS 314
RESULT 13
C47455
hypothetical protein T26112.40 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47655
R:Manfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Jencke, K.; Mayer, K.F.
Submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24471
A:Accession: T47655
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2149 <MON>
A:Cross-references: EMBL:AL132954
A:Experimental source: cultivar Columbia; BAC clone T26112
C:Genetics:
A:Map position: 3
A:Introns: 882/1; 1043/3; 1133/2; 1320/2; 2069/1
A:Note: T26112.40
Query Match 11.7%; Score 73; DB 2; Length 2149;
Best Local Similarity 23.0%; Pred. No. 2e+02;
Matches 31; Conservative 27; Mismatches 55; Indels 22; Gaps 4;
QY 3 NSLKKERVEDNCKSGNR-----RGKPSSTVVR-TVTEVEDE--FFK 43
DB 1330 SELKAATDLDTSQSDSNLANLVHPSLWPTILLRLKPS--TASESGDDLDPEVFM 1387
QY 44 ILRRVHVATVAKVNGGVAEGELPSKKRQNLGRNSLDCNGVRDGEFDEINRVGLQ 103
DB 1388 FIMKCSOSNLRVRLASRALVGLVSNEXLQSLRLTASTLPSNAGOGSGFNHLGILLQ 1447

GenCore version 5.1.3
Copyright (c) 1995 - 2002 Compend Ltd.

MM nucleic - nucleic search, using sw model

Run on: December 4, 2002, 06:37:47 : Search time 53 Seconds
(without alignments)
8275.405 Million cell updates/sec

Title: US-09-733-685-3
Perfect score: 1700
Sequence: 1 tgggtttttattgataaca.....ttacgaattacgaacatat 1700

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

searched: 441362 seqs, 15333381 residues

total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 94
Maximum Match 100%
Listing first 45 summaries

Database : issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCIRUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103.2	6.1	7218	1	US-08-232-463-14 Sequence 14, Appl
2	75	4.5	19124	2	US-08-487-826B-13 Sequence 13, Appl
3	71.2	4.2	152331	3	US-09-128-555-16 Sequence 15, Appl
4	71.2	4.2	176373	3	US-09-128-155-17 Sequence 12, Appl
5	62.6	3.7	152331	3	US-09-128-555-16 Sequence 16, Appl
6	58	3.4	6124	4	US-08-213-119B-3 Sequence 3, Appl
7	56.8	3.3	289	4	US-09-057-005-17 Sequence 17, Appl
8	56.8	3.3	289	4	US-09-244-796-17 Sequence 17, Appl
9	55.6	3.3	291	1	US-07-922-723A-7 Sequence 7, Appl
10	55.6	3.3	291	1	US-07-799-828C-7 Sequence 7, Appl
11	55.6	3.3	291	1	US-08-074-275-7 Sequence 7, Appl
12	55.6	3.3	291	1	US-08-486-366-7 Sequence 7, Appl
13	55.6	3.3	291	2	US-07-552-277A-7 Sequence 7, Appl
14	55.6	3.3	19124	2	US-08-487-826B-13 Sequence 13, Appl
15	55.6	3.3	44453	4	US-09-146-653-5 Sequence 5, Appl
16	54.2	3.2	5852	1	US-07-867-106-2 Sequence 2, Appl
17	53.4	3.1	1956	4	US-08-555-906B-1 Sequence 1, Appl
18	52.8	3.1	7379	4	US-03-341-587-5 Sequence 5, Appl
19	52.4	3.1	1298	4	US-08-971-089-1 Sequence 1, Appl
20	51.8	3.0	454	2	US-08-623-906A-6 Sequence 6, Appl
21	51.6	3.0	72604	4	US-09-268-992-7 Sequence 7, Appl
22	51.6	3.0	72604	4	US-09-657-474-7 Sequence 7, Appl
23	50.4	3.0	1559	4	US-09-018-095A-7 Sequence 7, Appl
24	50.2	3.0	2792	4	US-08-570-367C-1 Sequence 1, Appl
25	50	2.9	1298	3	US-08-948-705-3 Sequence 3, Appl
26	49.6	2.9	417	4	US-08-559-896B-3 Sequence 3, Appl
27	49.6	2.9	1859	3	US-08-691-563C-46 Sequence 46, Appl

Sequence 5, Appl
Sequence 4, Appl
Sequence 103, App
Sequence 103, App
Sequence 103, App
Sequence 10, Appl
Sequence 13, Appl
Sequence 595, App
Sequence 7, Appl
Sequence 7, Appl
Sequence 14, Appl
Sequence 1, Appl
Sequence 13, Appl
Sequence 6, Appl
Sequence 13, Appl
Sequence 14, Appl
Sequence 15, Appl
Sequence 16, Appl
Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-06-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/POCKET NUMBER: 30472/-14 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pl-Fls
US-08-232-463-14

Query Match 6.1% Score 103.2; DB 1: Length 7218;

JS-09-128-155-16

Query Match: 4.2%; Score 71.2; DB 3; Length 152331;
Best Local Similarity 54.6%; Pred. No. 1.2e-06;
Matches 142; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 203 AATAAAGATGAGATGAGATGAGATTTCTGTGTAAGAAAGAGAGAGATGATGAGTG 262
DB 128751 AGCAAGAAAG 128654
QY 263 AACAAATTAAGAGATATCATGATATATTTATGAGAGGTCGTGAAGATTATTTAGAG 322
DB 128691 AGAAAG 128632
QY 323 AGGAG 382
DB 128631 AGAAAG 128572
QY 383 AAGATGAG 442
DB 128571 AAG 128512
QY 443 ACAAAG 462
DB 128512 AGCAAG 128492

RESULT 4

JS-09-128-155-17/c
Sequence 17, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:

APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128.155
EARLIER FILING DATE: 1996-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 176373
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3)...(176373)
OTHER INFORMATION: n = A,T,C or G

JS-09-128-155-17

Query Match: 4.2%; Score 71.2; DB 3; Length 176373;
Best Local Similarity 54.6%; Pred. No. 1.3e-06;
Matches 142; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 203 AATAAAGATGAGATGAGATGAGATTTCTGTGTAAGAAAGAGAGAGATGAGTG 262
DB 144774 AGCAAGAAAG 144715
QY 263 AACAAATTAAGAGATATCATGATATATTTATGAGAGGTCGTGAAGATTATTTAGAG 322
DB 144714 AGAAAG 144655
QY 323 AGGAG 382
DB 144654 AGAAAG 144595
QY 383 AAGATGAG 442
DB 144594 AAG 144535

QY 443 ACAAAG 462
DB 144534 AGGAG 144515

RESULT 5

US-09-128-155-16
Sequence 16, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128.155
CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(152331)
OTHER INFORMATION: n = A,T,C or G

Query Match: 3.7%; Score 62.6; DB 3; Length 152331;
Best Local Similarity 51.2%; Pred. No. 0.0001;
Matches 146; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 173 ACATTGACGGAATTCAGGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 232
DB 146914 AGATGGAG 146973
QY 233 ATGGAAG 292
DB 146974 AGAAG 147033
QY 293 TATGAGAGAGTGGTGAAGATTATTTAGGAGAGGAGAGAGAGATAGAGAGAGAGAGAG 352
DB 147034 AAGGAG 147095
QY 353 ATGCTGAATCTGAAGAGAGATGAATTTGTTTAAAGATGAAGAGAGAGAGAGAGAGAG 412
DB 147094 AAG 147153
QY 413 CTAAAGTCTCTGAAGAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 457
DB 147154 AAG 147198

RESULT 6

US-09-128-155-17
Sequence 17, Application US/08213429B
Patent No. 633406
GENERAL INFORMATION:
APPLICANT: Inselburg, J. et al.
TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
FILE REFERENCE: J11-002CNC
CURRENT APPLICATION NUMBER: US/08/213.419B
CURRENT FILING DATE: 1994-03-14
PRIOR APPLICATION NUMBER: US 07/870,506
PRIOR FILING DATE: 1992-04-17
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0

```

SEQ ID NO 3
LENGTH: 6124
TYPE: DNA
ORGANISM: Plasmodium falciparum
FEATURE:
NAME/KEY: CDS
LOCATION: (2407)..(2439)
NAME/KEY: CDS
LOCATION: (2598)..(3404)
NAME/KEY: CDS
LOCATION: (3580)..(3720)
NAME/KEY: CDS
LOCATION: (3850)..(5835)
3-08-213-419B-3

Query Match      3.4%; Score 58; DB 4; Length 6124;
Best Local Similarity 46.%; Pred. No. 0.00044;
Matches 215; Conservative C; Mismatches 240; Indels 3; Gaps 1;

107 AAAATACATCGATTAAAAATCAGACATCTGGTTTAAATCAAAATCTAATCTCTTTATC 166
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2169 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2226
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 ACAAGCAGATTCAGCGAATTCAGGTAAAAAGAGAAATTAAGCAATGAGAGATAGAG 226
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2229 ATAAATTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2268
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 ATTCTATGAGAAAAGAGAGAGACATGTAGGTGACAAATTAACACATATGATGAT 286
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2289 AATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2348
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 ATATTTTATGAGAGGTGGTGAAGATTAATTTAGSAGAGGAGAGAGAAATAGAAAA 346
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2349 ATATTCAAAAAAATTAAGTCTCTAAATATTAATAACTTAAATTCTCAATATATCA 2408
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 AATGACATGCGSANTCTGAGAGAGATGANTCGTGTTAAGATGAGAGACAGAAAGAC 406
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2409 GAAGTCATATATTCCTGTGTTTCATATGTTGTAAGATGAAAAAANAANAAGAGGA 468
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
407 CCATGGCTAAAGTCTCTGAAGAGATGAAAAAGAAACAAAGAGAGAGAGAGAGAGA 466
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2459 AAGGAAAGAGAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2528
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
467 AAGGCTAAATAGACTTAATTCGCAAAATTTCTGACCGACAGAAATATATTGGTCC 526
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2529 ATAATATATATATAATAATAAT---ATAATATTTTACGCATATACAAACATTTC 585
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 AAGGTATTTTTGTATCTCTTTTGAAGTCAAAAGTTAT 564
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2586 TATTTTITTTTGTCTTATATTACAAAAATGTTAT 2525
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
S-09-007-005-17
Sequence 17, Application US/0907005B
Patent No. 6254558
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rihong
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
FILE OF INVENTION: FUSIONS
FILE REFERENCE: 00786/350003
CURRENT APPLICATION NUMBER: US/09/007,005B
CURRENT FILING DATE: 1998-01-14
EARLIER APPLICATION NUMBER: 60/035,963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064,491
EARLIER FILING DATE: 1997-11-06
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289

```

```

; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-C9-007-005-17

Query Match          3.3%;   Score 56.8;   DB 4;   Length 289;
Best Local Similarity 12.2%;   Pred. No. 0.00035;
Matches 31; Conservative 105; Mismatches 118; Indels 0; Gaps

QY 195 AAAGAGAAAATAAAGAAATGAGAGATAGAGACATTCTATCGAAAGAAGAGAGACA 254
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 33 RANGCARARURKARCRAURGRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 92
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 255 TGTAGGTGAACAAATAAGAGATATGATATATTTTATGAGAGTGGTGAAGATTAT 314
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 93 KNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 152
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 315 TTTAGCAGGGCAGAGAAATAGAAAAAGAAATGCATGGTCGAATCTCAAGAAGATCA 374
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 153 FNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 212
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 375 ATTGTGTAAAGATGAAGACAGAGAACATCCATGGCTAAAGTCTCGTAAGAAGATG 434
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 213 FKRNRSRCRANGRCRURGCRGRURARCRURGRURGRURGRURGRURGRUR 272
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 435 AAAAAGAAACAAAA 446
      |||||
Db 273 AAAAAAAAAAAAAA 286

RESID: 8
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rife
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: PJSIONS
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,492
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match          3.3%;   Score 56.8;   DB 4;   Length 289;
Best Local Similarity 12.2%;   Pred. No. 0.00035;
Matches 31; Conservative 105; Mismatches 118; Indels 0; Gaps

QY 195 AAAGAGAAAATAAAGAAATGAGAGATAGAGACATTCTATCGAAAGAAGAGAGACA 254

```


